MicroRNAs and target genes involved in *E. globulus* xylogenesis: *in silico* prediction and experimental validation

Clara Susana Marques Graça

Dissertation to obtain the Master Degree in

Biologia Funcional

Supervisor: Doctor Jorge Almiro Pinto Paiva

Co-supervisor: Doctor Leonor Morais Cecilio

Examination Commitee:

Chairperson: Professora Doutora Sara Barros Queiroz Amâncio, Professora Associada com Agregação do Instituto Superior de Agronomia da Universidade de Lisboa.

Members: Doutor Jorge Almiro Barceló Caldeira Pinto Paiva, Investigador Auxiliar do Instituto de Investigação Científica Tropical, I.P.

Doutor Nelson José Madeira Saibo, Investigador Auxiliar do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa.

Lisboa, 2013

Tese de Mestrado desenvolvida no âmbito do Projecto MicroEgo (PTDC/AGR-GPL/098179/2008), designado "Pediu algo pequeno? O poder dos microRNAs no mundo de *Eucalyptus* em tensão!", e do Projecto TreeForJoules (P-KBBE/AGR-GPL/0001/2010), designado "Improving eucalypt and poplar wood properties for bioenergy", atribuidos pela Fundação para a Ciência e Tecnologia (FCT).

FCT
Fundação para a Ciência e a Tecnologia

Agradecimentos

O trabalho apresentado nesta tese de Mestrado do Curso de Biologia Funcional foi desenvolvido no Laboratório de Células Vegetais, no Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa/Instituto de Biologia Experimental e Tecnológica, sob supervisão do Doutor Jorge Pinto Paiva.

À Fundação para a Ciência e Tecnologia (FCT), agradeço a concessão das Bolsa de Investigação no âmbito do Projecto MicroEgo (PTDC/AGR-GPL/098179/2008), designado "Pediu algo pequeno? O poder dos microRNAs no mundo de *Eucalyptus* em tensão!", e do Projecto TreeForJoules (P-KBBE/AGR-GPL/0001/2010), designado "Improving eucalypt and poplar wood properties for bioenergy".

Ao Doutor Jorge Pinto Paiva, agradeço por ter aceite ser meu supervisor e a oportunidade de entrar no mundo da investigação, através da aceitação da minha candidatura às Bolsas de Investigação no qual estou inserida. Agradeço a atenção prestada, os inúmeros desafios lançados e a partilha de conhecimentos. Agradeço também a amizade, a confiança, a (muita) paciência e o (enormíssimo) rigor necessários para que este trabalho pudesse ter algum sucesso. Também agradeço as caixas de *eppendorfs* cor-de-rosa lindas de morrer que me fizeram trabalhar com muito entusiasmo, as canetas dos congressos e todos os presentes que deu para me incentivar a trabalhar mais e melhor. Um Obrigado sincero por me dizer "Tu consegues Clara!" ou "Parabéns Clara!" (",).

À Professora Doutora Leonor Morais, agradeço por ter aceite ser minha co-orientado neste trabalho. Agradeço a atenção prestada a todas as questões, aos problemas com o meu (péssimo) inglês, e a todas as correcções que conseguiu fazer em tempo record.

Ao Professor Doutor Pedro Fevereiro, responsável pelo Laboratório de Biotecnologia de Células Vegetais (BCV), agradeço a oportunidade e o voto de confiança em mim depositado, aquando da aceitação da minha entrada para o laboratório, correndo um (sério) risco de me ter a trabalhar neste local. Obrigado "pelas pancadas" que me deu..pelos "puxões de orelhas"..pela amizade que no fundo (bem lá no fundo) eu sei que o professor tem por mim (",).

Agradeço também a todas as pessoas e instituições que participaram no desenvolvimento deste trabalho, nomeadamente: Doutor Laurent Farinelli (FASTERIS SA, Switzerland) pela preparação e produção das bibliotecas de smallRNAs e mRNAs; aos inúmeros parceiros nacionais e internacionais pela colaboração na área da bioinformática, Professora Ana Teresa Freitas, Doutor Nuno Mendes e Mestre Jorge Oliveira (INESC-ID, Portugal), Doutora Hélene San-Clement (UMR5546, França) e Doutor Georgios Pappas (Cenargen-EMBRAPA, Brasil); na área da biologia molecular, Doutora Jacqueline Grima-Pettenati e Doutor Marçal Soler (UMR5546); e na área do melhoramento florestal e instalação dos ensaios de indução de lenho de tensão, Engenheira Clara Araújo e Doutora Lucinda Neves (ALTRI FLORESTAL SA, Portugal).

Bem, parece que chegou a vez dos meus coleguinhas do BCV. Agradeço a todos os que de algum modo me ajudaram a desenvolver este trabalho ou simplesmente (com muito sacrifício, eu sei!!!) me

viram a brincar, a rir, a discutir e até a chorar... Muito Obrigada. Em especial não posso deixar de agradecer a algumas pessoas:

À "minha GRANDE equipa": Jorge Paiva (Chefe), Victor Carocha (Bituru), Susana Pera (Perinha), Jorge Oliveira (Junior) e Priscila (Pi). A todos agradeço a ajuda incansável prestada ao longo de todo este tempo (acabou pessoal, acabou!! weeeeeeeee), os conselhos do tipo "calma, Clara, tens de ter calma" e os ensinamentos importantíssimos ao longo de todas as actividades desenvolvidas. Um OBRIGADO gigante para vocês todos (",).

- Agradeço ao Victor pela enorme paciência que sempre (às vezes) teve comigo, ajudando-me e evoluir e a desempenhar as funções o melhor possível. Obrigada pela ajuda na compreensão dos conceitos, técnicas e processos que caíam nas minhas mãos como bombas e que só o Victor conseguia fazer com que de "bichos de sete cabeças" passassem a coisas simples e fáceis (Pois pois, fáceis né Victor... Tá bem abelha!! ... Abelha não, Carocha!!). Agradeço também o ensaio do Northern blot que desenvolveu e a preciosa ajuda no RACE que muito contribuiu para este trabalho.

- À Perinha agradeço pelas conversas, brincadeira, e gargalhadas que demos. Obrigado pela "Ideia do buraco" que fez com que tivessemos uma "tarde clandestina" e uma noite maravilhosa na praia ao som do Dj Diego Miranda, a dançar de dedos no ar e copos na mão!! Enfim, acabaste a noite a dormir em "casa da bruxa"...tiveste sorte que deixei a vassoura em casa senão podia ter sido mau...muito mau (",) Obrigada por teres "passado por cima do que a bruxa te fez"..e teres percebido (ou não) a pessoa que sou (Txiii..também não exageremos)!!

- Agradeço ao Junior por ter sido a primeira pessoa a reconhecer todo o meu talento e conhecimento, tratando-me por "Doutora" todos os dias, o que fazia com que os meus dias fossem substancialmente mais alegres (",).

- Por fim, mas não menos importante agradeço á Pi por me aturar e por não se chatear com as minhas brincadeiras...lembrando-a todos os dias que devido ao seu grande tamanho, 1,49 m no B.I. (grande precisão Pi, muito bem!) parece "um porta-chaves" e que ainda tem de crescer muito pa ser como eu (temo que não seja possível cresceres até ao 1,72 m, mas podes continuar a tentar (",). Também não me posso esquecer da expressão "Oh quê!" que me tirou tantas vezes do sério.

À Rita Morgado, agradeço pelos cafés e pelas pausas para "matar o vício". E, em especial, por conseguir dividir casa comigo e pela substituição na limpeza da casa quando eu estava cheia de trabalho e não pude cumprir com as minhas obrigações domésticas. Agradeço também o facto de me deixar a trabalhar sozinha pela noite dentro e não me fazer companhia. Ah, mas emprestando-me sempre o computador dela para que eu não passasse uma noite inteira a escrever no meu micro-magalhães (",). Espera lá...mas isto não aconteceu agora!! Já não divides casa comigo, já não me substituis na limpeza e não me emprestaste o teu pc. Esquece este agradecimento..Afinal não tenho que te agradecer!!! Ahhhh..agora assustaste-te =) bem, de facto já não "estamos juntas" nem na mesma casa nem no mesmo laboratório ☹ mas guardei todos os momentos...rimos muitas vezes juntas..chorámos também..partilhámos um bocado das nossas vidas é verdade.. há bocado disseste "esta miuda é fixe" e eu retribuo...Também és muito fixe e cinco estrelas =) Só é pena eu esperar por

ti todos os dias que acabes de almoçar e tu não esperares por mim para lanchar porque encontras "alguém" e foges =) Mas Obrigada.

À Maria Assunção ("a grávida") agradeço pelas gargalhadas, pela boa disposição e pelas brincadeiras. Obrigada por me responderes aquelas perguntas =) E não te esqueças de dizer ao Tomás que a "tia" não é má para a mãe =) .. Mas isto é só agora porque estás num estado de graça... porque daqui a uns meses as coisas vão mudar..Não vai puder continuar a ser tudo ao jeito da "grávida" =). À Mara Lisa, à Susana Leitão, ao Nuno Almeida, ao Marco Dinis, à Diana Branco, ao Zé Ricardo e aos restantes elementos deste enorme grupo pelas dicas, pela ajuda nas técnicas laboratoriais e no funcionamento do laboratório. Esperem lá..mas isto já foi há muito tempo...pronto..talves agora vos tenha de agradecer por outras coisas: Mara obrigado por me teres inspirado com as "cenas várias e diversas".. ficou muito bem no texto ☺; Susana Leitão obrigado por me ouvires a resmungar e calmamente ma pedires para "ter calma"..obrigado também por me lembrares que "só acabou por agora..daqui a um mês há mais",Nuno obrigado pelos cafés que nunca foste buscar à máquina nas horas de almoço..e pelas tuas "variadas e diversas" teorias tão tão...possas..faltam-me as palavras..enriquecedoras será??!! E, por fim, Zé Ricardo...bem tenho que te agradecer as pausas para os cafés..para os cigarros...para conversar...para rir...para chorar...Obrigada pelas expressões "bem vinda" ou "bem haja"..Obrigada pela dica dos agradecimentos "Clara escreves Um bem haja a todos. Obrigada. E pões um coração enorme cor-de-rosa a ocupar o resto da folha"...uaaaauuuuuuuuu..que belas dicas..vai lá ver a última frase =) !! Bem..obrigada pela tua amizade, pela paciência e pela ajuda com a leitura da tese, as formatações, o inglês...tudo.. Desculpa lá se te chateei muito!! Agradeço sinceramente a TODOS a disponibilidade, simpatia e o bom ambiente criado no laboratório. Ahhh... E agora "Tudo de castigo..de castigo"..Obrigado Rita Caré por nos pores todas na linha ☺

Aos meus familiares..sem dúvida os mais especiais:

- Aos meus pais, aos meus irmãos, aos meus cunhados. Em especial agradecer à minha irmã...oh irmã, desculpa lá as birras que faço, as guerras.. Agradeço-te por tudo aquilo que já fizeste por mim (ambas sabemos do que estou a falar)...deste-me sempre tudo o que precisei..deste-me carinho (txxiii..assério que estou a dizer isto?????), tecto, atenção, agasalho.. mostraste-me que havia esperança quando eu estava totalmente perdida ☺!! Adoro-te muito e agradeço-te o facto de me teres dado a Felicidade de ser tia =) agradeço também à minha pequenina Dianinha todas as alegrias que me dá e pelas birras que tantas vezes faz =) Ah, agradeço também ao cunhado por me ligar e dizer "Olá tas boa?", "Sim tou e tu?", "Também. Então se estás bem Adeus". Obrigada e Adeus (",).

- À verdadeira razão de isto tudo ter sido possível.. sem dúvida a pessoa mais importante... o Nandinho!! Obrigada do fundo do coração pelo amor, carinho e por me apoiar SEMPRE. Obrigado por todas as palavras e por me incentivar a ultrapassar mais este obstáculo na minha vida. Agradeço pelo esforço de no início "Tomar vir a Lisboa quando Lisboa não podia ir a Tomar" e agora no fim "Lisboa ir à Suiça quando a Suiça não podia vir a Lisboa". Agradeço todo o apoio sem nunca me deixar para trás um único segundo, apesar dos 2500 km que nos separam. Muito Obrigada AMT.

☺ A Todos, Um Bem Haja!! ☺

Abstract

Portugal is one of the largest producers of pulp and paper derived from *Eucalyptus globulus*, which makes this a valuable species for the country. Wood is a complex and variable material, and its post-transcriptional regulation knowledge is only beginning. MicroRNAs (miRNA) are small size (21-24nt), endogenous non-coding RNAs, involved in post-transcriptional regulation. *MiRBase* v20 database encloses thousands of entries, however none from *Eucalyptus*. In this study we aim to validate *E. globulus* miRNAs candidates; to predict *in silico* and validate experimentally the miRNAs targets; and analyze the gene expression of validated targets.

Four miRCa-02, miRCa-04, miRCa-08 and miRCa-09 candidates were validated by Northern blot and there *in silico* prediction revealed 42 target genes. Fourteen predicted target genes were tested through the RLM 5'-RACE methodology, but only three predicted targets were validated (Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113 predicted target genes for miR171, miRCa-04 and miRCa-08, respectively). Expression of these three target genes analyzed by RT-qPCR suggests that the distinct expression levels found may be related with to wood formation in *Eucalyptus globulus*.

For the first time, four *Eucalytus* miRNAs and their target genes were disclosed and validated by bioinformatic and molecular tools.

Keywords: *Eucalyptus globulus*, wood formation, miRNAs, target genes, RLM 5'-RACE.

Resumo

Portugal é dos maiores produtores de derivados de pasta e papel de *Eucalyptus globulus*, tornando esta espécie muito valiosa para o país. A madeira é um material complexo e variável, e o conhecimento da sua regulação pós-transcripcional está ainda no começo. MicroRNAs (miRNA) são RNAs endógenos não-codificantes de pequeno tamanho (21-24nt) que estão envolvidos na regulação pós-transcricional. A base de dados *miRBase v20* engloba milhares de entradas, no entanto nenhuma é de *Eucalyptus*. Neste estudo procedeu-se à validação de miRNAs candidatos de *E. globulus*; à predição *in silico* e validação experimental dos seus alvos; e à analise de expressão génica dos alvos validados.

Os candidatos miRCa-02, miRCa-04, miRCa-08 e miRCa-09 foram validados por *Northern blot* e a sua predição *in silico* revelou 42 genes alvo. Catorze genes alvo predictos foram testados através da metodologia RLM 5'-RACE, mas apenas três alvos foram validados (genes Eucgr.E01509, Eucgr.C01382 e Eucgr.J02113 alvos de miR171, miRCa-04 e miRCa-08, respectivamente). A expressão destes três genes alvo, analisada por RT-qPCR, sugere que os diferentes níveis de expressão encontrados estão relacionados com a formação da madeira em *Eucalyptus globulus*.

Pela primeira vez quatro miRNAs de *Eucalyptus* e os seus genes alvo são divulgados e validados através de métodos moleculares e bioinformáticos.

Palavras-chave: *Eucalyptus globulus*, wood formation, miRNAs, target genes, RLM 5'-RACE.

Resumo alargado

Portugal é um dos maiores produtores e exportadores de pasta e papel proveniente da madeira de *Eucalyptus globulus*, sendo esta uma das espécies florestais de maior interesse económico para o país. Este interesse deve-se sobretudo ao rápido crescimento desta espécie e à qualidade das suas fibras para a produção de pasta e papel.

A madeira é um material altamente complexo e variável, o que é particularmente visivel em certos tipos especiais de madeira formada sob condições contrastantes e estímulos ambientais: i) variação durante a estação de crescimento, com formação de madeira sazonal, ii) variação entre a madeira juvenil e adulta, e iii) variação em resposta à indução de um estímulo gravitrópico, com formação de madeira de reacção. A madeira sazonal e a madeira de reacção são consideradas excelentes modelos para o estudo dos mecanismos de regulação génica da xilogénese. Os microRNAs (miRNAs) são pequenas moléculas de RNA (20-24 nucleótidos) endógenas e não codificantes, que podem desempenhar importantes papéis de regulação, visando a clivagem ou a repressão da tradução do RNA mensageiro. A base de dados *miRBase v20* reúne mais de 24.500 miRNAs, porém nenhum diz respeito ao género *Eucalyptus*.

Este estudo teve como principal objectivo contribuir para o conhecimento dos papéis dos miRNAs na regulação pós-transcricional da xilogénese em *E. globulus*, particularmente no seu envolvimento na resposta a stresses ambientais e mecânicos. De modo a atingir o objectivo proposto procedeu-se: i) à validação de quatro miRNAs candidatos, identificados por ferramentas bioinformáticas; ii) à predição *in silico* dos genes alvo dos quatro miRNAs candidatos; iii) à validação experimental dos genes alvo predictos; e iv) à avaliação do perfil de expressão dos genes alvo validados, usando os dois modelos em estudo de formação de madeira sazonal e de formação de madeira de reacção, e outros tecidos e órgãos provenientes de *E. globulus*.

Numa primeira fase, quatro miRNAs candidatos foram selecionados para este estudo, a partir de uma lista de 3.300 miRNAs candidatos gerada no âmbito do projecto microEGo: miRCa-02, miRCa-04, miRCa-08 e miRCa-09, pertencentes a quatro familias MIR anteriormente identificadas noutras espécies por homologia de sequência: *MIR167*, *MIR396*, *MIR172* e *MIR477*, respectivamente. As formas guide e star* dos candidatos selecionados foram utilizadas para validar o seu tamanho e analisar o seu perfil de expressão por Northern blot nos modelos estudados. Posteriormente, com o intuito de identificar e compreender o modo de acção destes miRNAs, foi efectuada a predição *in silico* dos genes alvo para os quatro miRNAs candidatos selecionados (forma guide), usando o software *psRNATarget*. Adicionalmente, esta predição foi também efectuada para o candidato homólogo miR171, selecionado para testar a viabilidade da metodologia experimental de validação dos genes alvo. Seguidamente foi efectuada a validação experimental dos genes alvo dos miRNAs por RLM 5'-RACE, usando o modelo de estudo de formação de madeira de reacção. Para além da validação dos alvos, esta técnica permite detectar o local de clivagem que, geralmente ocorre entre o décimo e o décimo-primeiro nucleótido, a partir da extremidade 5' do miRNA. Tendo em conta os resultados obtidos na predição dos genes alvo e a sua potencial função no contexto da xilogénese, foram selecionados três genes alvo para cada miRNA candidato. Finalmente, foi realizada a análise

do perfil de expressão dos alvos validados, utilizando o RT-qPCR e os dois modelos de estudos de formação de madeira sazonal e de formação de madeira de reacção.

Ambas as formas dos quatro miRNAs candidatos apresentaram uma banda com o tamanho esperado (20-24 nt), pelo que se considerou que as predições foram efectuadas com sucesso. Por outro lado, os perfis de expressão obtidos revelaram diferenças entre os candidatos, com algumas variações de intensidade, sugerindo o seu envolvimento na diferenciação dos diferentes tipos de tecidos e do seu potencial papel na regulação da xilogénese em E. globulus. As predições dos genes alvo in silico permitiram identificar 68 transcriptos codificados por 42 genes alvo. Quarenta e oito por cento dos genes alvo predictos foram identificados como factores de transcrição e 43% foram associados a processos metabólicos e celulares, com funções de stress/defesa. Além destes, 9% dos genes alvo predictos aparecem com função desconhecida, sugerindo novos possíveis modos de acção dos miRNAs nesta espécie. Em 97% das predições, a clivagem do gene alvo mediada pelo miRNA é o mecanismo de regulação prevalente, e apenas 3% correspondentem à repressão da tradução. De todos os genes alvo predictos, foram testados experimentalmente quatorze por RLM 5'-RACE. De acordo com os resultados obtidos relativamente ao tamanho esperado dos produtos de amplificação, apenas os genes alvo Eucgr.E01509, Eucgr.E01875, Eucgr.B02279, Eucgr.C01382 e Eucgr.J02113 foram selecionados para prosseguir na validação experimental. Após a clonagem e sequenciação dos fragmentos, o local de clivagem esperado foi validado para os genes Eucgr.E01509, Eucgr.C01382 e Eucgr.J02113 alvos dos candidatos miR171, miRCa-04 e miRC-08, respectivamente. Algumas razões podem ser apontadas para este resultado, tais como i) as diferenças entre os genomas de *E. grandis* (640 Mbp, utilizado para a predição de alvos) e de *E. globulus* (530 Mbp, utilizado para a validação de alvos), ou ii) a possibilidade de existência de baixa especificidade dos iniciadores de PCR utilizados. A análise de expressão por RT-qPCR revelou que todos os genes alvo apresentam variações de expressão em ambos os modelos testados, existindo no entanto casos em que não existem diferenças estatisticamente significativas durante a formação de madeira sazonal e de madeira de reacção.

Esta foi a primeira vez que quatro candidatos miRNAs e alguns dos seus genes alvo em eucalipto foram estudados e validados por ferramentas moleculares e de bioinformática. Este trabalho contribuiu para a obtenção de um importante conjunto de recursos e para o conhecimento sobre os miRNAs e os genes alvo envolvidos na xilogénese de *E. globulus*. Os resultados aqui apresentados enfatizam a necessidade de se proceder à caracterização funcional dos miRNAs, como complemento aos estudos de sequenciação e análise dos perfis de expressão, de forma a compreender a sua função nas plantas.

Palavras-chave: *Eucalyptus globulus*, formação de madeira, microRNAs, genes alvo, RLM 5'-RACE.

Contents

List of Figures

List of Tables

Abbreviations

AdL – Adult leaves
AGO1 – Argonaute 1
A. thaliana – *Arabidopsis thaliana*
ARF – Auxin transcription factor
AP2 – APETALA2-like family
BAC – Bacterial artificial chromosome
BLASTn – Nucleotide basic local alignment search tool
cDNA – Complementary strand DNA
Ct – Threshold cycle
CTAB – Cetyl trimethylammonium bromide
CompW – Compression wood
CW – Control wood
DCL – Dicer-like protein
DCL1 – Dicer-like 1
Dec – December
DNA – Deoxyribonucleic acid
dNTP – Deoxyribonucleotides triphosphates
dsRNA – Double-strand RNA
DX – Developing xylem
E - Efficiency
EDC – 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA – Ethylenediamine tetraacetic acid
ESTs – Expressed sequence tags
E. globulus – *Eucalyptus globulus*
E. grandis – *Eucalyptus grandis*
EW – Early wood
Feb - February
G – Guaiacyl
GC – Guacine-cytosine
G-layer – Gelationous layer
GRF – Growth regulating factor
GSP – Gene specific primer
HEN – Hua enhancer 1 protein
HST – Hasty
HYL1 – Hyponastic leaves 1 protein
JW – Juvenile wood
LNA – Locked nucleic acid
LW – late wood

MM – Molecular marker
MFA – Microfibril angle
MIR – microRNAs genes
miRCa – MicroRNA candidate
miRNA – MicroRNA
miRNA* - MicroRNA star (complementary strand miRNA)
mM - Millimolar
MOPS – 3-morpholinopropane-1-sulfonic acid
mRNA – Messenger RNA
MW – Mature wood
NaCl – Sodium chloride
ng - Nanogram
nt – Nucleotide
NW – Normal wood
ncRNA – Non-coding RNA
ORF – Open reading frame
Ova - Ovaries
OW – Opposite wood
PAZ – Piwi Argonaute and Zwille
PCR – Polymerase chain reaction
Phl – Phloem
Pis – Pistils
Pol – Pollen
Pol II – Polymerase II
PPD – PAZ and PIWI domains
Pre-miRNA – microRNA precursor
Pri-miRNA – Primary miRNA
PVP – Polyvinylpyrrolidone
P. trichocarpa – *Populus trichocarpa*
QTLs – Quantitative trait *loci*
RFU – Relative fluorescence units
RG – Reference genes
RISC – RNA-induced silencing complex
RNA – Ribonucleic acid
rRNA – Ribosomal RNA
RLM 5'-RACE – RNA ligase-mediated 5' rapid amplification of cDNA ends
RT – Reverse transcription
RT-qPCR – Quantitative real-time PCR
Roo - Roots
RW – Reaction wood

RW_{2009} – Reaction wood campaign 2009
RW_{2010} – Reaction wood campaign 2010
RW_{2011} – Reaction wood campaign 2011
S – Syringyl
SCL – SCARECROW-like family
SD – Standard deviation
SDS – Sodium dodecyl sulfate
SE – Serrate protein
Sep – September
sRNA – Small RNA
siRNA – Small interfering RNA
snRNP - Small nuclear ribonucleoprotein
Sta – Stamens
SW – Seasonal wood
S/G ratio – Ratio syringyl/guaiacyl
tasiRNA – *trans*-acting small interfering RNA
TBE – Tris-Borate-EDTA
TE – Tris-EDTA
TF – Transcription factors
TG – target genes
Tris-HCl – Tris(hydroxymethyl)amino methane
TW – Tension wood
UTR - Untranslated region
UV – Ultraviolet light
V. vinifera – *Vitis vinifera*
YoL – Young leaves
µg - microgram

1

Introduction

Contents

1.1. *Eucalyptus* genus and *E. globulus* species

Forests are one of the most important natural resources of Europe and the plantations of *Eucalyptus* are grown in most continents with more than 20 million hectares planted worldwide (Mizrachi *et al*., 2010). In Portugal, forests occupy about thirty-nine percent of the national territory, occupying about 3.5 million hectares. In this area, sixty-seven percent (approximately 2.3 million hectares) are occupied by three main species considered to be major and also with greater economic interest, such as maritime pine (*Pinus pinaster*) with 885.0 hectares, eucalyptus (*Eucalyptus spp*.) with 740.0 hectares and cork oak (*Quercus suber*) with 716.0 hectares (Figure 1). Despite maritime pine being the species that occupies the largest forest area, *Eucalyptus* is also widespread and a very important genus in economical and ecological terms and therefore an inevitable presence in the Portuguese landscape (CELPA, 2011).

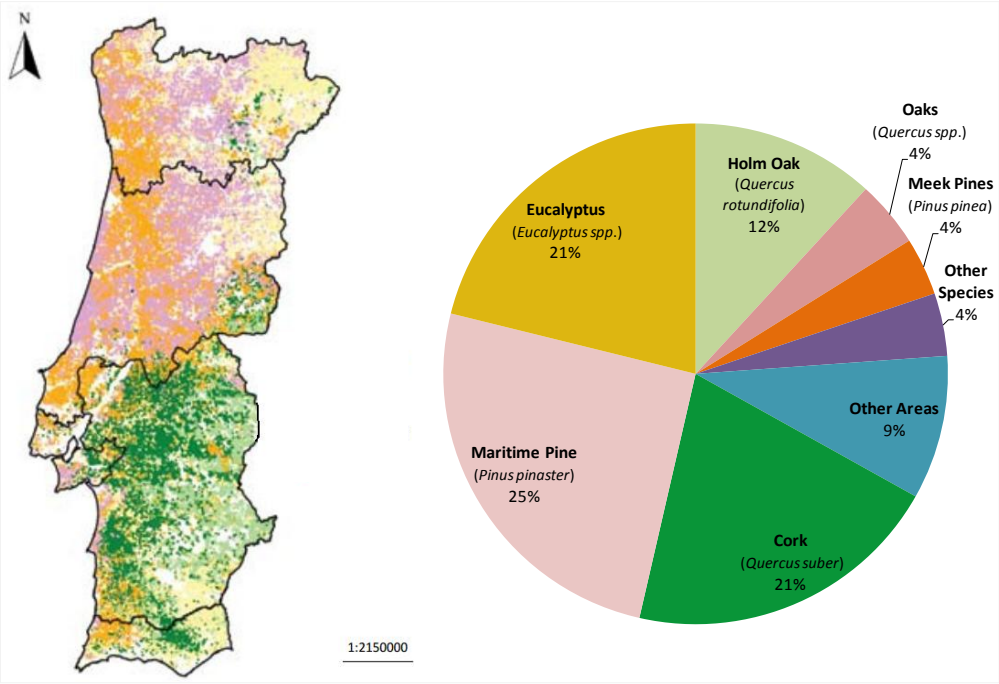


Figure 1: Geographic and spatial distribution of forest area in Portugal mainland (Source: adapted from CELPA, 2011).

The genus *Eucalyptus* was initially named by L'Héritier de Brutelle (1789) with a single species then, *E. oblique* (Brooker, 2000). It is an Angiosperm Dicotyledonous, belonging to the *Myrtales* order and it is a member of the *Myrtaceae* family. This genus includes some of the tree species which are most used in industrial forests plantation, since they are easy to grow and present wood with high density and durability (Plomion *et al*., 2001; Villar *et al*., 2011). These properties coupled with high forest productivities make them particularly suitable for the sectors of pulp and paper production, and for solid wood derived products (Potts *et al*., 2004; Pereira, 2007; CELPA, 2011). Certain eucalyptus species present some

of the tallest trees in the world, with large and fast growing, reaching over 70 meters tall (Pinho, 2010). Among the more than 700 described species, *Eucalyptus globulus* (*Labill.*) and their hybrids occupy the largest cultivated area in Portugal and is one of the most widely cultivated (Brooker, 2000; Villar *et al*., 2011). Its trunk is covered with a gray skin color and/or brown, smooth and hard bark that tend to drop spontaneously releasing the long strips that hanging, become dry, and persisting in the trunk for long periods. The juvenile leaves have a glauca color becoming green in adults, which are long, lanceolate with long petiole. The flowers are white with numerous stamens, and the fruits are woody capsules of green color (Figure 2) (Ladiges *et al*., 2003; Silva, 2007).

Figure 2: Different aspects of anatomical features of *Eucalyptus globulus*. (**a**) Adult tree; (**b**) Juvenile leaves glauca color; (**c**) Mature leaves of green color, whose branches have some fruit and flowers; (**d**) The open flower with numerous stamens; (**e**) Fruit/woody capsules green color; (**f**) Upper and lower capsules of the fruit separate; (**g**) Open capsule, the stamens are visible (left) located in the upper capsule, and the pistil (right) located lower capsule and below the of the ovaries; (**h**) Ovaries located in the capsule bottom, below the pistil.

The ability of sustaining economically viable *E. globulus* commercial plantations is a great national economical advantage since its wood is the primary raw material for the Portuguese pulp industries and because this is the paradigm species for obtaining high pulp quality, environmentally cheaper, an effective alternative to bio-fuels as well as the biggest atmospheric carbon sequestering agent (Boudet *et al*., 2003). The pulp and paper sector of Portuguese presented good productivity and their contribution has been positive for wealth creation in Portugal (Pereira *et al*., 2010; CELPA, 2011). Portugal is the fourth biggest European pulp and the 11th largest European producer of paper and cardboard. Although, Portugal has some favorable edapho-climatic conditions, important limitations in terms of cost of labor, soil availability and heterogeneity, coupled with the increasing competitiveness of Southern hemisphere plantations, demand novel and educated strategies for increasing

productivity and wood quality (Pinho, 2010; Bedon *et al*., 2011). Given its recognized wide natural genetic variability there is a still huge potential for increasing yields and that is being pursued by advanced breeding programs in countries like Australia, Brazil, Chile, Spain and Portugal (Costa e Silva *et al*., 2004; Potts *et al*., 2004; Parsons *et al*., 2006; Pereira *et al*., 2010). As a result of a combination of breeding programs and improved techniques for improved rooting ability, certain plantations have recently been established with clones selected for their wood productivity (Grattapaglia and Kirst, 2008) as well as wood quality and end-use properties (Raymond, 2002; Costa e Silva *et al*., 2004).

1.1.1. Recent advances in *Eucalyptus* research

In the last decade, a series of projects based on genomics and transcriptomics have provided a huge leap in the knowledge of *Eucalyptus* trees. Many *Eucalyptus* species are gaining special importance as model species in particularly for studies focusing on the xylem differentiation phenomena, providing valuable information regarding the wood formation mechanisms, stress response and comparative genomics (Paux *et al*., 2005; Poke *et al*., 2005; Neale, 2007; Grattapaglia *et al*., 2009). Moreover, *E. globulus* and *E. grandis* have a relatively small genome, 530Mbp (Poke *et al*., 2005) and 640Mbp (Grattapaglia, 2008), respectively. Therefore, and despite the differences in size of these two genomes, the most important gene families are expected to be conserved, allowing the use of the available *E. grandis* genome for *E. globulus* studies (Steane *et al*., 2001; Marques *et al*., 2002; Poke *et al*., 2005; Grattapaglia, 2008; Grattapaglia and Kirst, 2008; Hamilton *et al*., 2009). A great number of genetic and genomic tools are available with genetic maps and QTLs (Grattapaglia, 2008; Gion *et al*., 2011), ESTs collections (Rengel *et al*., 2009), BAC libraries (Barrela, 2011; Paiva *et al*., 2011), and the release of the full genome annotated sequence now publicly available at *www.phytozome.net* (Myburg *et al*., 2013; submitted).

Transcript profiling has been used by several groups to identify differentially expressed genes of the *Eucalyptus* species (Foucart *et al*., 2006; Salazar *et al*., 2013). Novaes *et al*. (2008) used 454 pyrosequencing technologies to discover a set of unigene sequences, providing a much needed public resource for *Eucalyptus* research, including the annotation of the *E. grandis* genome sequence. On the other hand, Cassan-Wang *et al*. (2012) established a set of 17 genes suitable for being used as a reference in High-throughput RT-qPCR analysis of expression in 13 different tissues and organs of *Eucalyptus*, in several experimental conditions.

1.2. Wood formation in trees

Among the multiple adaptations developed by land plants during its evolution, the acquisition of a vascular system, at approximately 400 million years ago, was a critical event for their successful colonization, adaptation and maintenance on Earth (Paux *et al*., 2004). The cyclic successive additions of secondary xylem, differentiated from the activity of the vascular cambium, give rise to a unique complex tissue called wood (Chaffey, 2002). Wood, also known as secondary xylem, is a hard fibrous tissue present in some plants to confer a stable structure, also serving as a pipeline for water, nutrients, hormones, and minerals allow them to rise from above ground. It is a composed tissue made of the successive stacking of growth rings, year after year, that provides mechanical support to the plant body that are xylem functions. Secondary walls are particularly important in these specialized cells that require strong walls to function (Ruiz-Medrano *et al.*, 2001; Pilate *et al*., 2004a). It is a complex and highly variable tissue given the heterogeneity of their cell types which present complex anatomic, chemical, and physical properties determined under strict ontogenic and environmental control. These wood properties derive from the size, shape, and arrangement of different cell types as well as the structure and chemical composition of the secondary cell walls of distinct xylem cells (Goicoechea *et al*., 2005).

Wood formation (xylogenesis) is a particular example of the complex developmental process known involving the differentiation of a meristematic tissue into the specialized cell types found in wood (Déjardin *et al*., 2010). This process occurs at the vascular cambium, a secondary meristem in the vascular tissue of the plants, composed of a lively population of cells able to divide in three planes (tangential, transversal and radial) and to differentiate into a variety of cell types comprising two different tissues: i) xylem to the adaxial (inner) side of the trunk and ii) phloem to the abaxial (outer) side. Wood is formed upon the death of xylem cells, and as a consequence of annual growth cycles, new layers of xylem are added to the existing wood around the trunk just under the bark (Merchan *et al*., 2009; Déjardin *et al*., 2010). During the differentiation of cambium in xylem cells, there is a dynamic and partial time and special overlapping phase: first a phase of radial expansion and intrusive growth followed by secondary wall deposition, and finally a programmed cell death (Han, 2001; Mellerowicz *et al*., 2001).

The cell wall is composed of several layers that are formed at different times during cell differentiation, where all the xylem cells undergo cell expansion until the deposition of secondary wall inside the primary cell wall after the primary wall has stopped growing (Mellerowicz *et al*., 2001; Plomion *et al.,* 2001). The first layer of cells that forms the middle lamella is found among the wood cells, and ensures adherence between neighbor cells (Chaffey, 2002). Then, there is the formation of the primary cell wall composed of several

layers of randomly oriented microfibrils. Between these microfibril layers we can found peptic substances, hemicelluloses and lignin (Darley *et al*., 2001; Plomion *et al*., 2001). At the time the cell reaches its definitive development, there is the formation of a new layer within their cell wall called the secondary cell wall (Chaffey *et al*., 2001). This layer is considered as the most important region for the cell in mechanical terms, and is divided into three different layers (S_1, S_2 and S_3) composed of cellulose microfibrils, which are highly oriented and in which we can also found hemicelluloses and lignin (Plomion *et al*., 2001; Figure 3). This differentiation into specialized types of cells is a highly ordered cellular mechanism that involves an extremely sophisticated order of events, which are under temporal and spatial regulation in order to achieve the coordination of various molecular signals and the differential expression of many genes (Jones, 2001; Ranik *et al*., 2006).

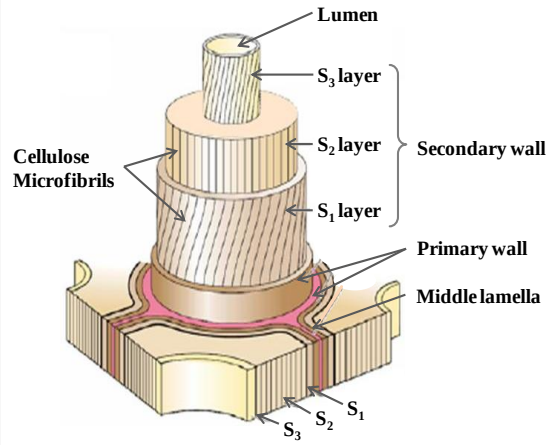


Figure 3: Three-dimensional structure of the cell wall organization often found in cells with thick secondary wall. The cell wall is divided into several layers (primary wall, S_1, S_2 and S_3 layers), each with their own arrangement and orientation of cellulose microfibrils (Source: Nuno Borralho from IICT).

1.3. Chemical composition of cell wall

One of the main issues concerning wood formation is the stage of secondary cell wall thickening, which is also one of main factors that determine the composition and wood structure. Secondary cell walls are mainly composed of cellulose (40-50%), lignin (25-35%), hemicelluloses (20-30%) and proteins (Christensen *et al*., 2000). Table 1 shows the estimation of wood chemical composition for five *Eucalyptus* species (Magaton *et al*., 2000).

Table 1: Chemical composition and S/G ratio of some of the main commercial *Eucalyptus* species (Source: Magaton *et al*., 2000).

Species	Cellulose (%)	Hemicelluloses (%)	Total lignin (%)
E. globulus	48.8	24.5	26.7
E. grandis	51.7	20.5	27.8
E. nitens	43.8	30.9	25.3
E. urograndis	47.5	24.5	28.0
E. urophylla	52.0	19.4	28.6

1.3.1. Cellulose composition

Cellulose is the most abundant natural origin polymer (biopolymer) in the planet. It occurs naturally in the form of fibers that are composed of interlocking chains of β-1,4-glucan (Barnett and Bonham, 2004). In cell wall, the spaces between the cellulose microfibrils are occupied by hemicelluloses and lignin, and the deposition of cellulose occurs outside the cell membrane and differs significantly in two major cell types. The primary cell walls are composed of thin layers of flexible and randomly deposited cellulose microfibrils during cell differentiation (Mellerowicz *et al*., 2001). During the later stages of maturation, the secondary wall is deposited between the primary cell wall and cell membrane. The cellulose microfibrils of the secondary cell wall are largely crystalline and are arranged in very specific angles that differ significantly not only between species, but also between different wood types (e.g. normal and reaction wood) (Barnett and Bonham, 2004).

1.3.2. Hemicelluloses composition

The hemicelluloses are polysaccharides composed of non-identical monomers of carbohydrates with low molecular weight, associated with cellulose in plant tissues. These polysaccharides generally occur as heteropolymers, and it's fibers become closely associated with water insoluble cellulose microfibrils, which are associated with mixtures of soluble non-cellulosic polysaccharides (Plomion *et al*., 2001).

1.3.3. Lignin composition

Lignin is a phenolic polymer derived from three hydroxycinnamyl alcohols; giving rise to the two constitutive units guaiacyl (G) and syringyl (S), which only differ among them in their degree of methoxylation (Rodrigues *et al*., 2000; Boerjan *et al*., 2003). Lignin incorporates a matrix of polysaccharides, which confer stiffness and cohesion to the whole wood tissue, providing a hydrophobic surface required for the transport of water. The content and monomeric composition of lignin varies significantly among different *taxa*, individuals, tissues, cell types and cell wall layers (Grima-Pettenati *et al*., 1999; Christensen *et al*., 2000).

1.4. Wood variability

Wood is a highly complex and variable material. This variability is visible particularly in certain special types of wood formed under contrasting conditions and environmental stimulus such as: i) variation during the growing season (within the annual growth rings), ii) variation between juvenile and mature wood, and iii) variation during the response to gravitropic stimulus (between the normal and reaction wood) (Plomion *et al*., 2001; Ballarin and Palma, 2003).

1.4.1. Seasonal effect in xylogenesis

The seasonal wood (SW) effect is one of the most significant environmental sources of variation affecting cambial activity and the formation of new cells, therefore influencing wood quality (Zobel and Buijtenen, 1989). In temperate regions, the annual course of cambial activity (dormancy and activation) is activated by temperature and/or photoperiod (Uggla *et al*., 2001; Paiva *et al*., 2008). As the name suggests, the early wood (EW) formation occurs in the spring (beginning of the growing season), when temperature and photoperiod are favorable for vegetative growth and when the cambium is more active presenting cells with wider and thinner walls (Arend and Fromm, 2007; Déjardin *et al*., 2010). On the other side, in late summer or early autumn (end of the growing season), occurs formation of the late wood (LW), when delayed as the cambium is less active presenting cells with narrower and thicker walls (Paiva *et al*., 2008). Curiously, the transition from one type of wood to another could be abrupt between two growing seasons, i.e. between the LW and the following EW, but from EW and the following LW. This transition might be gradual or abrupt depending on the species (Pallardy, 2008).

1.4.2. Juvenile and mature effect in xylogenesis

The juvenile wood (JW) and mature wood (MW) effect occurs during all tree life, through the formation of growth rings. The juvenile wood is a major source of variability of the wood inside the tree. During the first years of activity, the vascular cambium produces wood with different characteristics from those that will produce later on (Ballarin and Palma, 2003; Arend and Fromm, 2007). At this time, during young stages of vascular cambium is produced the JW and later is produced the so-called MW. The JW always occupies the central part of the trunk. Since the top of the tree always has a young stage of cambium, this wood type has always being produced at the treetop regardless of the tree age (Roberts and McCann, 2000). Therefore, the juvenile and mature wood is produced simultaneously in the trees, with the MW in stem base after cambium maturation and the JW in the top. During the period of

transition from juvenile to mature wood, which occurs approximately 5 to 20 years, the characteristics of wood produced tend to improve gradually to become relatively constant, being this wood known as mature wood (Roberts and McCann, 2000; Leonello *et al*., 2008).

1.4.3. Gravitropic effect in xylogenesis

Other external influences like gravitropic stimulus result in reaction wood (RW) formed in response to a non-vertical orientation of the stem, caused either by a mechanical or by an environmental force that can be both naturally or artificially applied (Paux *et al*., 2005; Paiva *et al*., 2008). This response is based on the development of a particular type of wood, with special mechanical properties, allowing the tree to regain its equilibrium (Déjardin *et al*., 2010). The characteristics and location of this wood type differ between Gymnosperms and Angiosperms trees. In Gymnosperms trees (e.g. *Pines*), the reaction wood is named compression wood (CompW) and develops at the lower side of leaning stem and branches, and it contains a higher lignin content, as well as a lower amount of cellulose (Plomion *et al*., 2000; Paiva *et al*., 2008). In Angiosperms trees (e.g. *Eucalyptus*), reaction wood is formed on the upper side of the leaning stem and is named tension wood (TW). The wood formed on opposite side of reaction wood, is named opposite wood (OW). Compared to normal (vertical) wood (NW), tension wood shows higher tensile stress, fewer vessels, higher cellulose content and almost no lignin (Qiu *et al*., 2008), representing an excellent model for studying the formation of xylem cell walls (Figure 4; Pilate *et al*., 2004b). In particular, tension wood cells contain an extra layer, translucent and gelatinous-like cell wall layer (G-layer), in which results from the replacement of the secondary wall layer (S_3) by a layer slightly lignified, consisting almost entirely of highly crystalline cellulose (Jourez *et al*., 2001; Qiu *et al*., 2008). These modifications correspond to dramatic changes in chemical, physical and technological characteristics especially if compared with the ones presented by OW that shows characteristics far more similar to normal wood formed on a non-bended tree (Figure 4) (Pilate *et al*., 2004a). A high proportion of reaction wood in a trunk can be, in some cases, a major problem for several industrial applications. Reaction wood has several commercially undesirable characteristics such as a tendency toward shrinking, warping, weakness, and brittleness, which adversely affect its utilization, namely in the pulp production (Pilate *et al*., 2004a; Pallardy, 2008).

Figure 4: Induction of a gravitropic stimulus to stimulate tension wood formation in *Eucalyptus* trees. (**a**) Harvest local (open box) tension wood (TW) formation is induced on the upper side of bended trunk and opposite wood (OW) on the lower side. (**b**) Displays a detail of TW anatomy and (**c**) a detail of OW anatomy of the xylem cells stained with Safranin-Astrablue dyes: the lignified cell walls present a red color, and the G-layer (crystalline cellulose) an intense blue color (Source: Dr. Jorge Paiva and Teresa Quilhó from IICT).

1.5. Regulation of wood formation

Because wood is an extreme example of huge regulation and coordinated activation of specific genes, both seasonal wood (SW) and induction of reaction wood (RW) have been considered excellent models to study the impact of a one year growing time period over xylogenesis and to study xylem cell-wall biosynthesis (Bedon *et al*., 2007; Merchan *et al*., 2009). The wood formation is a notable example of cell differentiation in an exceptionally complex form, controlled by numerous gene families and an extreme plasticity of the involved metabolism, which adds further complexity that limits our understanding of this process (Plomion *et al*., 2001). After completion of the differentiation process and the formation of secondary xylem (wood) and mature phloem tissues, these tissues are functionally and chemically distinct. Xylogenesis requires the coordinate expression of a large number of the genes, some of which have been shown to be differentially expressed in the developing tissues surrounding the vascular cambium (Hertzberg *et al*., 2001; Burton *et al*., 2000; Ranik *et al*., 2006). Many genes and proteins involved in xylogenesis, including those of lignifications pathway (e.g. Cinnamoyl CoA reductase, *Egr*CCR) have been catalogued (Paux *et al*., 2005; Zhong and Ye, 2007; Brach *et al*., 2008; Novaes *et al*., 2008; Paiva *et al*.,

2008; Qiu *et al*., 2008; Rengel *et al*., 2009; Nascimento *et al*., 2011; Carocha *et al*., 2013), but in opposition, the network of regulatory mechanisms implied in their expression remains largely unknown. The study of the regulatory mechanisms involved in the wood formation is the basis for establishing the criteria for the selection of appropriate genotypes. Extensive studies have concentrated on the identification of gene families participating in this biological process. However, the identification of other major regulators, such as microRNAs, is still lacking (Zhong and Ye, 2007).

Transcription regulation represents major controlling steps determining tissue-specific and developmental stage-specific activity of many genes. Regulation by transcription factors (TFs) is an integral part of a highly complex network (Yamaguchi-Shinozaki and Shinozaki, 2005; Druart *et al*., 2007). The analysis of lignification genes has also shown their presence in the promoter of conserved motifs that have been demonstrated to be important in xylem localized gene expression (Carocha *et al*., 2013). Proteins that can bind this motif and activate the transcription belong to the MYB family. Two MYB genes preferentially expressed in *Eucalyptus* xylem were proposed to be involved in regulating transcription during xylogenesis (Goicoechea *et al*., 2005; Legay *et al*., 2007). Genetic mapping coupled with Quantitative trait loci (QTL) detection allowed the identification of several genomic regions associated with wood traits variation (e.g. Thamarus *et al*., 2004; Gion *et al*., 2011), such as wood density, pulp yield, or chemical composition (cellulose and lignin content, S/G ratio) leading to the identification/validation of candidate genes for these traits (Thamarus *et al*., 2004; Thumma *et al*., 2010; Gion *et al*., 2011; Mandrou *et al*., 2011). Other than transcriptional regulatory mechanisms and post-translational modifications, seem to play an important role in the regulation of gene expression (Hirayama and Shinozaki, 2010). An example of this post-transcriptional regulation are the small RNAs molecules.

1.6. Plant small RNAs

Evidence for the existence of RNA-mediated gene silencing (RNA silencing) mechanisms in plants first appeared in the late 1990's, when short antisense RNA molecules were isolated from tomato plants where post-transcriptional gene silencing had been detected (Hamilton and Baulcombe, 1999). Since then, the knowledge on non-coding small RNA (sRNA) has broadened and these molecules have been identified as important players in a wide variety of processes in plants, such as direct the sequence-specific down-regulation of target genes at the post-transcriptional level. This regulation can occur at some of the most important levels of genome function, including chromatin structure, chromosome segregation, transcription, RNA processing, RNA stability and translation (Mette *et al*., 2000; Bartel, 2004). Depending mainly on their origin, sRNAs can be roughly divided into two major

distinct classes: i) small interfering RNAs (siRNAs) that are processed from long, perfectly double-stranded RNA and ii) microRNAs (miRNAs) from single-stranded RNA transcripts (transcribed from MIR genes) that have the ability to fold back onto themselves to produce imperfectly double-stranded stem loop precursor structures (Doench *et al*., 2003; Rajagopalan *et al*., 2006; Eamens *et al*., 2008). These classes are the most broadly distributed in both phylogenetic and physiological terms and are characterized by the double-stranded nature of their precursors (Carthew *et al*., 2009).

MicroRNAs (miRNAs) are abundant classes of non-coding endogenous small RNAs usually with single-stranded molecules of 20-24 nucleotides (nt) (Bartel, 2004). In higher plants miRNAs act as negative post-transcriptional regulators, down-regulating mRNA with high sequence-specific (Ghildiyal and Zamore, 2009; Merchan *et al*., 2009; Hirayama and Shinozaki, 2010). The multitude of small non-coding RNAs (ncRNAs) found in eukaryotes (siRNAs and miRNAs) introduces an additional complexity in our understanding of the mechanism and proper classification of miRNAs (Bartel, 2004). The methods for identifying plant miRNAs are based on the major characteristics: (1) all miRNAs are endogenously expressed producing ~ 22 nt RNAs (Reinhart *et al*., 2002; Bartel, 2004); (2) all miRNAs precursors have a well-predicted stem loop hairpin structure, and were potentially processed from one arm of fold-back precursors (Reinhart *et al*., 2002; Bonnet *et al*., 2004), (3) many miRNAs are generally evolutionarily conserved (Berezikov *et al*., 2006), and (4) originate from regions of the genome distinct from previously annotated genes (Bartel, 2004).

1.6.1. Biogenesis of plant microRNAs

The broad concept of miRNA biogenesis requires multiples steps in order to form mature miRNAs from MIR genes, such as described in the Figure 5 (Bartel, 2004; Kurihara and Watanabe, 2004; Brodersen and Voinnet, 2006). First, a miRNAs gene (MIR) located generally in intergenic regions (Llave *et al*., 2002b; Reinhart *et al*., 2002) is transcribed to a primary miRNA (pri-miRNA), which is usually a long sequence of more than several hundred nucleotides (Jones-Rhoades and Bartel, 2004; Xie *et al*., 2005; Voinnet, 2009). This step is controlled by RNA polymerase II (Pol II) enzyme (Bartel, 2004; Kurihara and Watanabe, 2004; Lee *et al*., 2004) and produces a 5'-capped and 3'-polyadenylated primary transcripts, in a way similar to protein-coding genes (Xie *et al*., 2005). Second, the pri-miRNA is cleaved to a stem loop intermediate called miRNA precursor or pre-miRNA double-stranded RNA molecule (Kurihara el al., 2004; Lee *et al*., 2004; Kurihara *et al*., 2006). In plants, this step is controlled by the Dicer-like 1 (DCL1) enzyme, specifically involved in miRNA accumulation, and responsible for both cleavage steps in the nucleus (Bernstein *et al*., 2001; Papp *et al*., 2003; Tang *et al*., 2003; Meyers *et al*., 2008).

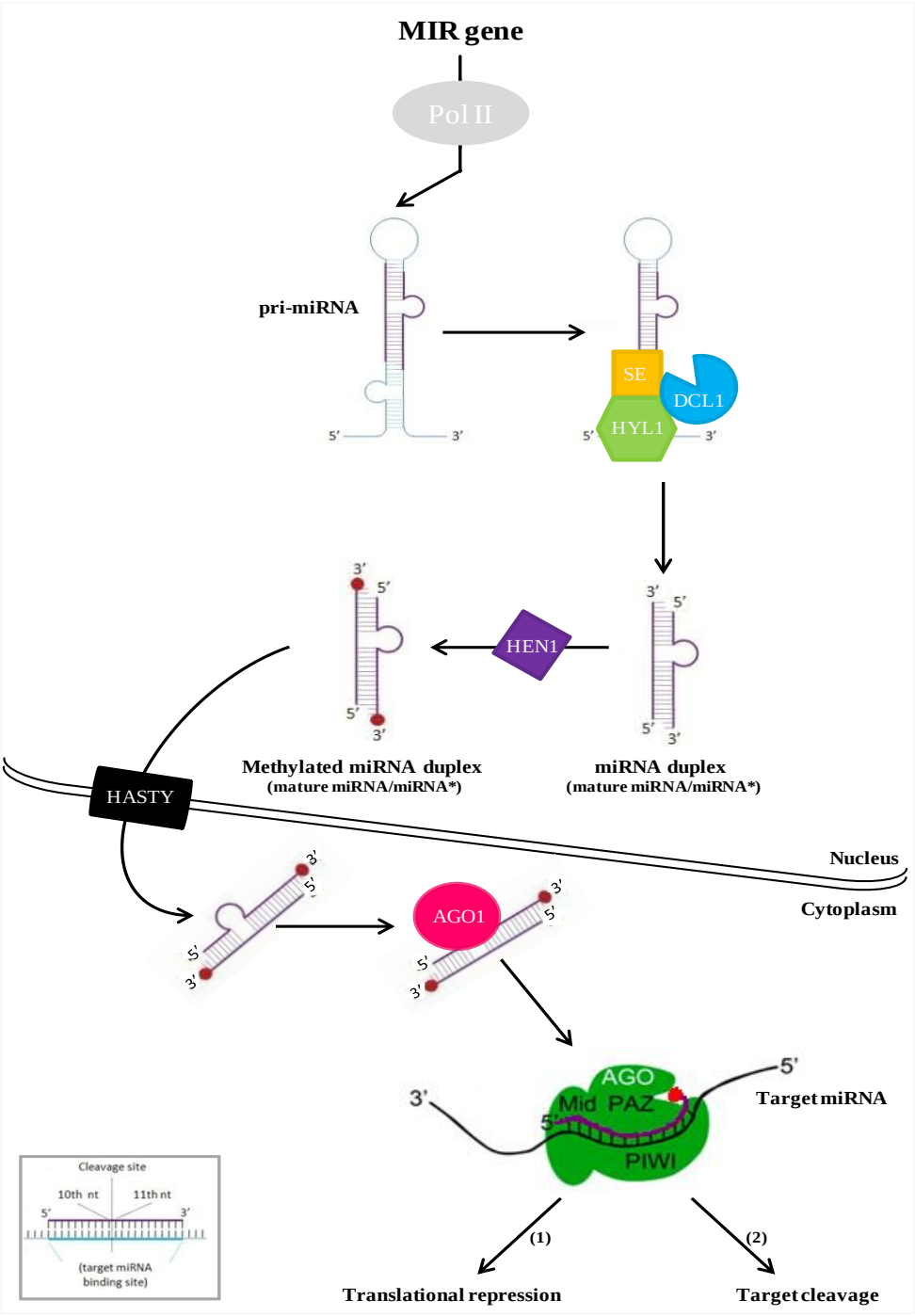


Figure 5: Simplified schematic representation of miRNA maturation and target cleavage in plants. Plant MIR genes are transcribed by RNA Polymerase II (Pol II). Transcripts fold into a hairpin structure denominated as primary miRNA (pri-miRNA). The DCL1, HYL1 and SE proteins form a complex DCL1 that cleaves it into a shorter precursor miRNA (pre-miRNA) dsRNA molecule. The DCL1 and HYL1 are involved in the second cleavage of the pre-miRNA into the mature miRNA. The 3' ends of miRNAs and the complementary strands (miRNA*) are methylated by HEN1, which protects against small RNA degrading enzymes. The nuclear exporter HASTY exports the mature miRNAs to the cytoplasm where one strand of miRNA duplexes associates with AGO 1 proteins integrated in the RISC complex, and (1) target translation repression, or (2) directs cleavage of partially complementary target mRNA. AGO proteins have three domains: Mid, PAZ and PIWI. It has been proposed that the Mid domain binds miRNA 5' ends while the PAZ domain binds miRNA 3' ends to position miRNAs for PIWI-mediated cleavage, of partially complementary mRNA target sites. The cleavage site is located between the nucleotides 10 and 11 starting from the 5' end of the miRNA sequence. DCL1 – Dicer-like 1; HYL – HYPONASTIC LEAVES1; SE – SERRATE; HEN1 – HUA ENHANCER 1; AGO1 – ARGONAUTE 1; RISC – RNA induced silencing complex (Source: adapted from Mallory *et al.*, 2008).

The pre-miRNA molecule folds to form an imperfect hairpin dsRNA stem-loop structure with the mature miRNA located on one arm (Lee and Ambros, 2001). Subsequently two other proteins, HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE), act in conjunction with DCL1 cleavage of this molecule on each arm forming the miRNA:miRNA* duplex, containing the mature miRNA and the complementary strand miRNA* (Reinhart *et al*., 2002; Kurihara *et al*., 2006), from the pre-miRNA stem-loop structure (Figure 5) (Lee *et al*., 2003; Han *et al*., 2004; Lobbes *et al*., 2006). The miRNA and the miRNA* remain together after cleavage, and are methylated by the sRNA-specific methyltransferase HUA ENHANCER 1 (HEN1) (Elbashir *et al*., 2001; Chen *et al*., 2002; Yu *et al*., 2005). The duplexes of siRNAs are also methylated by HEN1, a process that appears to be plant specific and are assumed to protect all sRNA species from degradation (Li *et al*., 2005; Yang *et al*., 2006).

At this point in the biosynthesis pathway, miRNA:miRNA* duplex can either stay in the nucleus or be transported into the cytoplasm through HASTY (HST) protein (Figure 5) (Bartel, 2004; Park *et al*., 2005). Finally, the mature miRNAs are then incorporated into a protein complex where regulate targeted gene expression named RNA-induced silencing complex (RISC) whose major protein component is an ARGONAUTE1 (AGO1) protein (Mi *et al*., 2008; Mallory *et al*., 2008; Schwarz *et al*., 2008), and act upon highly or perfectly complementary targets by promoting mRNA cleavage or repressing translation (Llave *et al*., 2002b; Rhoades *et al*., 2002). The AGO1 protein structure includes two conserved regions, the PAZ and PIWI domains (PPD) (Carmell *et al*., 2002). The PPD domain of AGO1 and the PAZ domain of Dicer may mediate Argonaute-Dicer interactions to allow for the transfer of the siRNA/miRNA molecule from Dicer to RISC (Yan *et al*., 2003; Farazi *et al*., 2008; Mallory *et al*., 2010). Most miRNAs are 21-nt long and are generally loaded into AGO1 and mediates miRNA-guided cleavage of complementary target transcripts (Figure 5) (Chen, 2004; Vaucheret *et al*., 2004; Baumberger and Baulcombe, 2005). All these steps suggest a complex process involving several enzymes required for processing long pri-miRNA to approximately 20-24 nt mature miRNAs (Chen *et al*., 2010; Cuperus *et al*., 2010; Shivaprasad *et al*., 2012).

1.6.2. Diversity of plant microRNAs functions

Plant miRNAs have a high degree of sequence complementarily to their target mRNAs and direct the cleavage of the target mRNAs in the middle of the complementary regions (Figure 5) (Llave *et al*., 2002a; Tang *et al*., 2003; Wang *et al*., 2005). It is observed that the perfect or near-perfect base pairing with the target RNA could promote cleavage of the target RNAs (Carthew and Sontheimer, 2009). The action of most plant miRNAs commonly involves a direct target cleavage/target degradation (Llave *et al*., 2002a, Rhoades *et al*.,

2002; Emergy *et al*., 2003; Kasschau *et al*., 2003; Palatnik *et al*., 2003; Tang *et al*., 2003; Achard *et al*., 2004; Juarez *et al*., 2004; Kidner and Martienssen, 2004; Mallory *et al*., 2004; Baker, 2005; Brodersen *et al*.,2008) or, in a few cases, translational repression, which affects the pool of transcripts remaining after cleavage. (Aukerman and Sakai, 2003; Chen, 2004). Importantly, neither the position (5'ORF or 3'UTR) nor the degree of pairing at the miRNA target sites appeared to be predictive of the prevalence of one process over the other (Voinnet, 2009). Although miRNAs are relatively small, they play an important role in gene expression, and the miRNAs are considered as one of the most important post-transcriptional gene regulators (Carrington and Ambros, 2003) as it was originally recognized in 2001 (Lee and Ambros, 2001; Lagos-Quintana *et al*., 2001). Several experiments have demonstrated that many miRNAs regulate various plant development processes, including leaf morphogenesis and polarity (Bowman *et al*., 2002; Emery *et al*., 2003; Juarez *et al*., 2004; Mallory *et al*., 2004; Kim *et al*., 2005), floral differentiation and development (Aukerman and Sakai, 2003; Achard *et al*., 2004; Chen, 2004), root initiation and development (Bartel and Bartel, 2003; Laufs *et al*., 2004; Mallory *et al*., 2005; Guo *et al*., 2005), vascular development (Lu *et al*., 2005; Kim *et al*., 2005), and induced for environmental and stress responses (Sunkar *et al*., 2004; Lu *et al*., 2005). The miRNAs are grouped into gene families, according to the similarity of the mature sequence. Certain families have a large number of members (e.g. miR156 or miR166), while others have just one or two (e.g. miR162 or miR390). However, this number can vary considerably among different plant species (Jones-Rhoades *et al*., 2006). Several studies describes an important set of these and others miRNAs functions in plants, presented in Table 2.

Table 2: Example list of miRNAs targets and their functions in plants (adapted from Zhang *et al*., 2006; Guo *et al*., 2005; Zhou *et al*., 2007; Rodriguez *et al*., 2010).

MiRNA	Mode of Action	Target protein class
miR156/157	Transcriptor factor	SQUAMOSA-promoter binding proteins
miR159	Hormone response	MYB
miR160/167	Signaling transduction	Auxin response factor
miR162	MiRNA biogenesis	Dicer-like protein
miR164	Plant morphogenesis	NAC domain transcription factors
miR165/166	Meristem formation and others	HD-ZIP transcription factors
miR166	Vascular development	HD-ZIP transcription factors
miR168	MiRNA functions	ARGONAUTE protein
miR171	Developmental patterning	SCARECROW-like (GRAS domain) transcription factors
miR172	Floral development and phage change	APETALA2-like transcription factors
miR173/miR390	siRNA biogenesis	Trans-acting siRNA
miR393/miR394	Hormone response and others	F-box protein, a class of bHLH transcription factors
miR395	Environmental stress response	ATP sulfurylase; sulfate transporter
miR396	Modulation of cell proliferation and expansion	Grouth regulating factors

A majority of these miRNAs affect plant traits by regulating the expression of TFs and to playing important roles in several aspects of plant development, such as perception and integration of cellular and environmental signals (Rhoades *et al*., 2002; Mallory *et al*., 2004; Mallory *et al*., 2005; Jones-Rhoades *et al*., 2006). Several families of TFs that depend on these conserved miRNAs to exert their roles on regulation of developmental functions are demonstred as follows: MYB domain proteins regulation of anther development by miR159 (Millar *et al*., 2005), NAC determination of organ boundaries by miR164 (Laufs *et al*., 2004), or HD-ZIP regulation of organ polarity and vascular arrangement by miR165 and miR166 (Williams *et al*., 2005; Zhou *et al*., 2007).

1.7. Identification of microRNAs in plants

The identification of miRNAs and their functions has opened new paths for an improved knowledge of the mechanisms of regulation of various development processes (Lu *et al*., 2005). The identification of miRNAs follows three main strategies: i) forward genetics, ii) bioinformatics prediction, and iii) direct cloning and sequencing (Chen, 2008). The first strategy, seeks to find the genetic basis of a phenotype. Despite having resulted in the identification of a few miRNAs, this method provides hints to the functions of these miRNAs in addition to their isolation (Chen, 2008).The second strategy for miRNA discovery is the bioinformatic prediction that employs several different algorithms to predict miRNAs (Adai *et al*., 2005; Mendes *et al*., 2009). The features that the algorithms seek in the genomic sequences are based on the search for homologous sequences of known miRNAs, and this can be developed both in a single genome of our interest or in genomes of related organisms. These features include the intergenic location of the MIR genes, the high degree of sequence complementarity of miRNAs to their mRNA targets, the hairpin structures of the precursors, and the conservation of some miRNAs between species (Szittya *et al*., 2008, Mendes *et al*., 2009). Finally, the third strategy for miRNA discovery is based on RNA sequencing after cloning and construction of libraries of small RNAs by a next generation sequencing techniques (Moxon *et al*., 2008). This strategy has the advantage of enabling the find of small non-conserved RNAs and species-specific miRNAs, but it presents also the disadvantage that the gene expression levels might be insufficient for the experimental detection, as miRNAs are temporal and might not be present in significant amounts in the sampled tissues (Unver *et al*., 2009).

1.8. Registry of microRNAs in *miRBase* database

The information about individual miRNAs from various organisms, including sequence data of precursors and mature forms, as well as genomic positions is accessible in a

database from the Sanger Institute: “The miRNA registry - *miRBase*” *http://www.MiRBase.org/* (Griffiths-Jones, 2004; Griffiths-Jones *et al*., 2006; Griffiths-Jones *et al*., 2008; Kozomara and Griffiths-Jones, 2011). This database includes several species grouped into miRNAs gene families containing a varying number of members, based on sequence similarity, where members differ from each other by a maximum of three nucleotides, target different members of the same gene families, and are derived from different genomic location, which expected to be regulate by the same or similar miRNAs (Rhoades *et al*., 2002; Dezulian *et al*., 2006).

In July 2013, the m*iRBase* release *version 20* (June 2013) has raised the total number of identified, published miRNAs to 24,521. This collection includes miRNAs of primates, rodents, birds, fish, flies, worms, plants and viruses. There are 6,145 miRNAs from *Viridiplantae* clade (*http://www.MiRBase.org/cgi-bin/browse.pl*). However, althrough miRNAs are registered for relevant plant species at the level of comparative genomics, such as *Pinus taeda*, *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera*, no records exist with regard to any of the three most important forest species for the portuguese economy, *Pinus pinaster*, *Quercus suber* e *Eucalyptus spp*. The Table 3 show the *MiRBase* information about the registration of woody plants miRNAs (*http://www.MiRBase.org/cgi-bin/browse.pl*).

Table 3: Number of sequences registries from several woody species from *Viridiplantae* Clade: *Coniferophyta* and *Magnoliophyta* Divisions, respectively (Source: *miRBase* database, v20 of June 2013).

Division	Family	Specie	Sequences registries	Reference
Coniferophyta	*Pinaceae*	*Picea abies*	40	Yakovlev *et al*. (2010)
		Pinus densata	30	Wan *et al*. (2012)
		Pinus taeda	36	Lu *et al*. (2007)
Magnoliophyta	*Brassicaceae*	*Arabidopsis thaliana*	298	Reinhart *et al*. (2002); Rhoades *et al*. (2002); Xie *et al*. (2005); Lu *et al*. (2006); Moldovan *et al*. (2010)
	Euphorbiaceae	*Ricinus communis*	63	Zeng *et al*., (2010)
	Rosaceae	*Malus domestica*	206	Arenas-H *et al*. (2009); Xia *et al*. (2012)
	Rutaceae	*Citrus clementine*	5	Song *et al*., (2009)
		Citrus reticulata	4	Song *et al*., (2009)
		Citrus sinensis	60	Xu *et al*. (2010)
		Citrus trifoliata	6	Song *et al*. (2009)
	Salicaceae	*Populus euphratica*	4	Li *et al*. (2009)
		Populus trichocarpa	352	Tuskan *et al*. (2006); Lu *et al*. (2005); Dezulian *et al*. (2006)
	Vitaceae	*Vitis vinifera*	163	Jaillon *et al*. (2007); Mica *et al*. (2009); Pantaleo *et al*. (2010)

1.9. Objectives

This work aims to provide new insights on the roles of small RNA-mediated regulation of gene expression during xylogenesis, specifically their role on the environmental and mechanical stresses responses in *Eucalyptus globulus*.

The first part of the work we aim to identify miRNAs that were differentially expressed under environmental and mechanical stresses. We will validate the selected miRNAs candidates identified *in silico* through Northern blot analysis using two models of wood formation and other tissues and organs from eucalypt. Subsequently, we aim to disclose the roles of the miRNAs candidates. Several miRNAs candidates will be selected and their targets will be predicted *in silico*. To validate experimentally the predicted targets we will use RLM 5'-RACE methodology. Finally, the expression profile of the validated miRNA target genes will be analyzed to provide hints about their role in xylogenesis.

The activities and results presented here were developed in the frame of the microEGo Project (Fundação para a Ciência e a Tecnologia funded PTDC/AGR-GPL/098179/2008, "Did you ask for something small? The microRNAs power in Eucalyptus tension world!") and TreeForJoules Project (Fundação para a Ciência e a Tecnologia funded P-KBBE/AGR-GPL/0001/2010, "Improving eucalypt and poplar wood properties for bioenergy").

2

Material and Methods

Contents

2.1. Plant material (*Eucalyptus globulus*)

2.1.1. Developing xylem sampling

The developing xylem (DX) samples were collected by sampling individual trees exposed to conditions inherent to our two models of study: seasonal wood (SW) and reaction wood induction (RW). Additionally, was collected developing xylem of mature trees (MW) with ten years old from a single genotype (VC9). After detaching the bark, the exposed soft gelatinous cell layers (2-3mm thick) corresponding to the DX tissue, were immediately scratched, using scalpel blades, into sterile flasks previously cooled with liquid nitrogen (Figure 6). All the collected tissues samples were kept away from defreeze and stored until use at -80ºC.

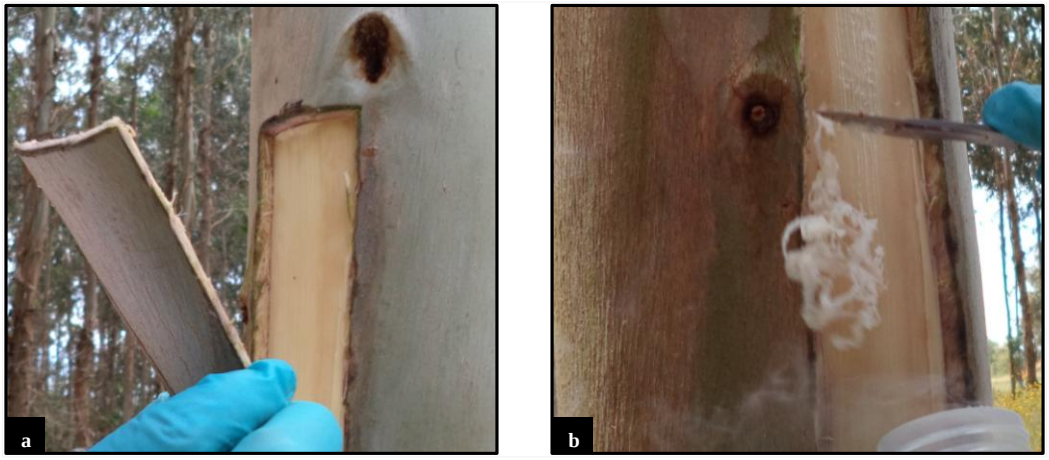


Figure 6: Harvesting of developing xylem tissue in *E. globulus*. (a) Removal of the tree bark. (b) Collection of developing xylem tissue by scraping it into a sterile bottle pre-cooled with liquid nitrogen.

2.1.1.1. Seasonal wood model

The sampling of seasonal wood (SW) was made in the year 2008, in collaboration with *RAIZ* Institute (Portugal). A population involving a total of thirty-six trees from three non-related distinct genotypes (HD161, CN5 and G-70) was sampled at Herdade do Zambujal, Pegões (Portugal), at breast height (1.30 m). All mature trees have been previously certified in terms of clonal identity by *RAIZ* using a battery of microsatellites as defined from their internal procedures. Four sampling campaigns were performed during a year: 26th February (Feb), 23rd May, 5th September (Sep) and 3rd December (Dec). For each condition and each genotype, three individual's biological replicates (ramets) were collected.

2.1.1.2. Reaction wood induction model

Three campaigns of tension (reaction) wood (RW) induction were conducted at Quinta do Furadouro, Óbidos (Portugal) in collaboration with our industrial partner *ALTRI FLORESTAL SA*: the first in 2009 (RW_{2009}), the second in 2010 (RW_{2010}) and the third in 2011 (RW_{2011}), using three individual's biological replicates (ramets) of three distinct genotypes (GM2-58, GB3 and MB43). The trees aged between four (RW_{2009}), five (RW_{2010}) and six (RW_{2011}) years old were bent with an angle of roughly 45º during different bending times. In the RW_{2009} campaign, a total of nine trees were bent at 11th July during three weeks, and harvested along with the non-bent trees (controls) on 31st July. In the RW_{2010} campaign, a total of forty five trees were bent in five different bending times: 9th June, 16th June, 23rd June, 30th June and 6th July, and harvested along with the non-bent trees (controls) on 7th July. Finally, in the RW_{2011} campaign, a total of thirty-six trees were bent at four different bending points: 24th June, 30th June, 15th June and 21st July, and harvested along the non-bent trees (controls) on 22nd July. For each campaign, DX were collected from each bent tree both tension wood (TW, upper side) and opposite wood (OW, under side). Additionally, in the same day, were collected the controls (CW) and three biological replicates for each bending time and each genotype. The time points made for each of the two campaigns are summarized in Figure 7.

	Bending Date					Sampling Date
RW_{2009}	-	11th July	-	-	-	31st July
RW_{2010}	9th June	16th June	23rd June	30th June	6th July	7th July
RW_{2011}	24th June	30th June	-	15th June	21st July	22nd July

Figure 7: Timeline representing the time of induction of gravitropic stimulus for the formation of *E. globulus* reaction wood (RW_{2009}, RW_{2010} and RW_{2011}) in Quinta do Furadouro, Óbidos (Portugal). Kinetic points sampled over four weeks (green) and date of collection of all points, including controls (yellow).

2.1.2. Others tissues and organs sampling

The sampling of other tissues and organs (T/O) was carried out in the year 2010, in collaboration with our industrial partner *ALTRI FLORESTAL SA* located in Quinta do Furadouro, Óbidos (Portugal). A total of seven types of *E. globulus* tissues were harvested in certified trees from four genotypes (GM2-58, GB3, MB43 and C33). Separate samples of

adult leaves (AdL), young leaves (YoL), pollen (Pol), roots (Roo), phloem (Phl) and floral tissues were harvested. The floral tissues such as stamens (Sta), ovaries (Ova) and pistils (Pis) were separated in a laboratory by dissection of the flowers.

2.2. Nucleic acids extraction

2.2.1. Total RNA extraction from *Eucalyptus globulus*

The total RNA extraction was performed with approximately 100 mg of biological material, for all samples described in subsection §2.1., according to a protocol adapted from Provost *et al*. (2007) with the following modifications: RNA extraction buffer (2% CTAB, 2% PVP, 100 mM Tris-HCl (pH 8.0), 25 mM EDTA, 2 M NaCl and 0.5 g/L Spermidine) was previously heated at 65^{o}C for 10 minutes and then added 2% β-mercaptoethanol. Grind tissue into a fine powder in liquid nitrogen using a mortar and pestle sterilized, and then mixed with 900 µL the buffer and incubate at 65^{o}C for 15 min. Extract three times with an equal volume of chloroform:isoamyl alcohol (24:1), separating phases by centrifugation at > 14,000 x*g* at room temperature for 15 min. Remove the final aqueous layer and precipitate the RNA in 0.1X of the total volume with 4 M sodium acetate (pH 5.2) and 2.5X of the total volume with absolute ethanol at -20^{o}C for 1 hour. After centrifugation at room temperature for 30 min, the RNA pellet was washed with 250 µL 70% ethanol and then with absolute ethanol. Finally, after discarding the ethanol, the pellet was dried and resuspended in 50 µL of *Milli*-Q RNase-free fresh water, and stored until use at -80ºC.

2.2.2. Total RNA integrity and quantification

The total RNA integrity was checked by running the newly extracted samples in a 2% (w/v) agarose gel in 0.5X Tris-Borate-EDTA buffer (TBE), prepared from the concentrated to 10X TBE solution (890 mM Tris-Base, 890 mM Boric acid, 20 mM EDTA), and stained with SYBR-Safe (Invitrogen®, Carlsbad, CA, USA). The total RNA was previously denatured by addition of formamide buffer (dionized formamide, 0.5 M EDTA pH 8.0, xylene cyanol and bromophenol blue) at 65ºC for 10 min. The molecular marker used was 1 Kb Plus DNA Ladder (Invitrogen®, Carlsbad, CA, USA), and the bands visualized in a *Gel Doc-1000 UV* image acquisition system, and analyzed in *Quality One* software (Bio-Rad® Laboratories, Inc.). The integrity and quality was checked taking in account the presence and sharpness of ribosomal RNA (rRNA) bands 28S and 18S.

The total RNA purity and quantity was assessed using the *NanoDrop ND-1000* Spectrophotometer (Thermo Scientific™, Wilmington, Delaware, USA), recording the total RNA concentration (ng.µL^{-1}). The optical density measured at different lengths allows the

assessment of nucleic acids purity. The absorption wavelength of contaminants and background is assumed to be at 230 nm, the absorption maximum for nucleic acids at 260 nm and the proteins at 280 nm. The ratio of absorption at 260 nm over 280 nm ($A_{260/280}$) is generally accepted to be a good indicator of RNA purity and its value should be approximately 2.00. If the ratio is appreciably lower, it may indicate the presence of protein, phenol or other contaminants that absorb strongly near 280 nm (Gallagher and Desjardins *et al*., 2006). The $A_{260/230}$ ratio was also measured and its value should be expected to range of approximately 2.00-2.20. If the ratio is appreciably lower than expected, it may represent the degree of background and contaminations (e.g. polysaccharides), which absorb at near 230 nm (Thermo Scientific™, Wilmington, Delaware, USA).

2.3. Prediction and identification of miRNAs candidates

2.3.1. SmallRNA-*Hiseq* libraries sequencing

In order to construct the smallRNA-*Hiseq* (non-coding) libraries, equimolar pools of total RNA were gathered according to the type of tissue and bending time of each sample, such as following described: i) for the seasonal wood (SW) model, four pools were constructed by DX sampling date corresponding to each season of the year, and ii) for the reaction wood (RW_{2010}) induction model, eleven pools were constructed by DX type (tension, opposite and control wood) corresponding to each bending and sampling time, respectively (see Table 4).

Table 4: Total RNA pools extracted from developing xylem tissues formed under seasonal wood and reaction wood models. These pools were sent for smallRNA-*Hiseq* libraries preparation and sequencing.

Plant Material		**Pool Code**
Seasonal Wood (SW)	February	Feb
	May	May
	September	Sep
	December	Dec
Reaction Wood (RW_{2010})	Control (non-bent)	CW
	4 weeks of bending	4OW
	4 weeks of bending	4TW
	3 weeks of bending	3OW
	3 weeks of bending	3TW
	2 weeks of bending	2OW
	2 weeks of bending	2TW
	1 week of bending	1OW
	1 week of bending	1TW
	1 day + 1 week of bending	1D1OW
	1 day + 1 week of bending	1D1TW

Legend: OW – opposite wood; TW – tension wood; CW – Control wood.

Ten micrograms of each total RNA pool were delivered to *FASTERIS SA* (Genève, Switzerland; *www.fasteris.com*) for libraries preparation and sequencing, using *FASTERIS SA* internal protocols. Before libraries preparation the quality of the samples were reanalyzed using the *Agilent 2100 Bioanalyzer* (Agilent Technologies®, Santa Clara, CA, USA). Values RNA integrity number (RIN) below seven is an indication of not appropriate RNA quality. The obtained smallRNA libraries were submitted to deep-sequencing through the sequencing service, using *HiSeq 2000* from *Illumina Genome Analyzer* service.

2.3.2. Bioinformatic identification of miRNAs candidates

After submission the results sequencing by *FASTERIS SA* company, the bioinformatics pipeline was used to analyze the *E. globulus* smallRNAs-*HiSeq* data generated by deep sequencing, according to the procedure described in Oliveira *et al*. (2013, accepted). The summarized steps as following described: i) only the sequence tags ranging between 18 nt to 30 nt in length were selected for further analysis; ii) the sequence tags holding homopolymers and unknown bases were discarded; iii) the sequences were trimmed whenever base quality score was lower than 15 using *INESC-ID* developed *PERL* scripts; iv) the remaining sequence tags still ranging between 18 nt to 30 nt in length were aligned to the *E. grandis* genome (*Phytozome v7.0*, annotation *v1.0*) allowing up to two mismatches using *Bowtie 0.12.7*; v) uniquely aligned sequence tags which were annotated by positional comparison as being *E. grandis* coding regions were also excluded using *BEDtools*. The remaining sequence tags were searched against: i) currently known miRNAs from model species *A. thaliana*, *V. vinifera* and *P. trichocarpa* (stored in *miRBase* v17); ii) the *A. thaliana* miRNAs database (*http://bioinformatics.cau.edu.cn/PMRD/*); iii) *A. thaliana* non coding RNAs; iv) *A. thaliana* repeats; and v) *E. grandis* repeat databank (containing all predicted *E. grandis* genome annotated repeated elements, developed and kindly provided by Doctor Georgios Pappas from *Cenargen-EMBRAPA*, Brasília, Brazil). After this exclusion stage, all the remaining predicted miRNAs were classified as homologs. *In silico* prediction of putative miRNAs in *E. grandis* genome (*Phytozome v7.0*, annotation v1.0, Goodstein *et al*., 2012) was made using the *CRAVELA* framework (Mendes *et al*., 2010; *www.cravela.org*) previously developed by *INESC-ID* for metazoans and adapted on purpose for the search of plant miRNAs. The *CRAVELA* predictions allowed to identify a list of 3,300 potential miRNAs candidates, using the smallRNA-*Hiseq* data analysis (Oliveira *et al*., 2013; accepted).

2.4. Northern blot validation of selected miRNAs candidates

Northern blot hybridizations were used for expression validation of eight potential mature miRNAs candidates (Table 5) predicted from four precursors (both guide and star forms).

These candidates were selected from the list of ten conserved eucalypt miRNAs that were found commonly (presenting full-sequence homology), between the set provided by *CRAVELA* predictions (Oliveira *et al*., 2013, accepted) and a eucalypt miRNAs independent database (Pappas *et al*., unpublished). These four selected miRNAs (guide form) are conserved candidates, representatives of four conserved MIR families previously identified in various species by sequence homology: miRCa-02 (MIR167), miRCa-04 (MIR396), miRCa-08 (MIR172) and miRCa-09 (MIR477) (Table 5). To gain support for validation and expression of these predicted miRNAs, the hybridizations were carried out using RNA samples from different tissues selected to cover a spectrum of potential miRNAs expression.

Table 5: List of four conserved eucalypt miRNAs selected from the list of ten conserved eucalypt miRNAs obtained based on the comparison of *CRAVELA* predictions and a eucalypt miRNAs independent database (kindly proved by Doctor Georgios Pappas, Brasília, Brazil). Also displayed are its homologs found in the *miRBase* v20 (Release 20.0, June 2013; Griffiths-Jones, 2004).

Precursor name	miRNA candidate mature sequence (both guide and star* forms)	miRNA size (nt)	Gene family	miRBase19 (1st hit homolog miRNA)
mirC-6538426	**miRCa-02**: UCAGAUCAUGCUGGCAGCUUCA	22	MIR167	ccl-miR167a
	miRCa-02*: UGAAACUGCCAGAUGAUCUGA	22		ath-miR167d
mirC-7274614	**miRCa-04**: CUUCCCACAGCUUUCUUGAAC	21	MIR396	cca-miR396a
	miRCa-04*: GUUCAAGAAAGCUGUGGAAG	20		zma-miR396g
mirC-8742859	**miRCa-08**: UGAGAAUCUUGAUGAUGCUGC	21	MIR172	vvi-miR172d
	miRCa-08*: GCAGCAUCAUCAAGAUUCACA	21		ptc-miR172i
mirC-8853075	**miRCa-09**: GAAGAACUUGGGGGAGUGCG	20	MIR477	aqc-miR477g
	miRCa-09*: CCUGCUCCCUCAAGGGCUUC	20		aqc-miR477e

Legend: * - star form; ath - *Arabidopsis thaliana*; ccl - *Citrus Clementine*; cca - *Cynara cardunculus*; zma - *Zea mays*; vvi - *Vitis vinifera*; ptc - *Populus trichocarpa*; aqc - *Aquilegia caerulea*.

2.4.1. Plant material

Two sets of tissue panels were defined to validate the expression for eventual differential presence of the miRNAs candidates in tissues generated under conditions relevant within the frame of the two xylogenesis models of study. A set of membranes was assembled for each panel, using RNA pools from the various genotypes and tissues in study (Table 6).

The first set, “developing xylem panel” (DX-panel) includes exclusively eight different types of xylem tissues. These samples comprised three sets: (1) four SW pools constituted by three distinct genotypes (HD161, CN5 and G-70), (2) three RW_{2009} pools constituted by three distinct genotypes (GM2-58, GB3 and MB43), and (3) one mature wood (MW) pool sampled in the DX of trees from a single genotype (VC9). These pools were defined under the assumption that different types of wood may increase chances of detecting expressed miRNAs assuming their active role in an ongoing, fully established *E*. *globulus* xylogenesis. The second set, “other tissues/organs panel” (T/O-panel) includes total RNA from nine different *E. globulus* tissues/organs, were also harvested in certified trees from four

genotypes (GM2-58, GB3, MB43 and C33). These samples will be used to confirm if the miRNAs candidates identified are indeed exclusively or preferentially expressed in developing xylem tissues.

Table 6: Total RNA pools defined for set up the several *E. globulus* samples to be included on the panels for Northern blot (one of the strategies used for the expressional validation of four miRNAs candidates).

Panels of tissues	Plant Material		Pool Code
Developing xylem panel (DX-panel)	Seasonal wood (SW)	February	Feb
		May	May
		September	Sep
		December	Dec
	Reaction wood (RW_{2009})	3 weeks of bending	09OW
		3 weeks of bending	09TW
		Control	09CW
	Mature wood (MW)	Mature	MW
Other tissues/organs panel (T/O-panel)	Other tissue and organs	Adult leaves	AdL
		Young leaves	YoL
		Roots	Roo
		Stamens	Sta
		Pistils	Pis
		Ovaries	Ova
		Pollen	Pol
		Phloem	Phl
		Developing xylem*	DX

Legend: OW – Opposite wood; TW – Tension wood; CW – Control wood. *Developing xylem obtained by the addition of equal quantity of each of the individual samples corresponding to seasonal wood (SW), reaction wood (RW_{2009}) and mature wood (MW).

2.4.2. Electrophoresis, blotting and cross-linking of RNA

The preparation and production of membranes was performed according to the protocol adapted from Trindade (2012). For each sample, around 25 µg of total RNA at a final volume of 20 µL was loaded and size separated in a 15% polyacrylamide gel (7 M urea in MOPS buffer), by electrophoresis on a *Mini-protean II Electrophoresis* device (Bio-Rad® Laboratories, Inc.), at 100V for 3-4 hour, in 1X MOPS buffer. The molecular weight marker used was microRNA marker, a mix synthetic RNA of 19 nt and 24 nt (Ambion®, New England Biolabs, Ipswich, USA). The gel separated RNA samples was transferred to a *Hybond-NX* membrane (GE Healthcare®, Piscataway, NJ, USA) by electro blotting with *Bio-Rad's Trans-Blot Semi-Dry* system (Bio-Rad® Laboratories, Inc.), at 20V for 1 hour in a 4ºC chamber. After the transfer, the hybridization membrane was then placed with the RNA side faced up on a *Whatman* paper moistened and chemically crosslinked with soluble EDC and 1-methylimidazole, according to Pall *et al*. (2007).

2.4.3. Hybridization

The LNA probes (Exiqon, Vedbaek, Denmark) were end-labeled with [γ^{32}P]-ATP (PerkinElmer®, Waltham, MA, USA) at 37ºC for 1 hour, using T4 polynucleotide kinase (Invitrogen®, Carlsbad, CA, USA). Labeled probes were purified with *Ilustra G-25 MicroSpin* columns (GE Healthcare®, Buckinghamshire, England). Hybridizations were carried out overnight at 42ºC for LNA probes in *ULTRAhyb-Oligo* buffer (Ambion®, Austin, TX, USA). Membranes were washed at hybridization temperature using a 2x SSC more 0.1% SDS (v/v) solution until radioactivity levels was reduced to less than 300 cpm. After 16 to 72 hours of exposure, depending on the radioactive signal intensity detected in the membranes with a Geiger counter, blots were visualized using a *PhosphoImager* system (Storm 860, GE Healthcare®, Buckinghamshire, England). To be re-used to a maximum of five times, the membranes were stripped by incubation in boiling 0.1% SDS solution, until the solution reached room temperature, and hybridized to the U6 probe sequence that is part of the non-coding small nuclear RNA component of U6 small nuclear ribonucleoprotein (snRNP), following the same protocol described above. The U6 allows for the quantification of all miRNA according to the U6 snRNP loading control, and therefore to signal normalization. The miRNA visible bands were quantified using the *ImageJ 1.45s* software (Schneider *et al*., 2012) in order to quantify the signal intensity for each spot. The measures were normalized against the corresponding signal intensity for U6 loading control among samples from the same membrane and the each specific hybridization. The values were exported in an *Excel* file and used to generate histograms.

2.5. Target genes prediction of selected miRNAs

The prediction of miRNAs target genes was performed to identify products of miRNA-directed cleavage for guide form sequence of four miRNAs candidates (miRCa-02, miRCa-04, miRCa-08 and miRCa-09) and one hyperconserved miRNA (miR171). The miR171 has been identified *in silico* in *Eucalyptus* miRNA dataset, and it was described in *Populus* trees as having interesting functions at the mechanical stress-responsive level (Lu *et al*., 2005). The *in silico* prediction of target genes were conducted with the *psRNATarget: A Plant Small RNA Target Analysis Server* (Dai and Zhao, 2011; *http://plantgrn.noble.org/psRNATarget/)* adopting the following set-up values: i) maximum expectation of 3.0, ii) length of complementarity scoring (hspsize) of 20 bp, iii) target accessibility - allowed maximum energy to unpair the target site (UPE) of 25, iv) flanking length around target site for target accessibility analysis of 17 bp in upstream and 13 bp in downstream, and v) range of central mismatch leading to translational inhibition of 9-11 nt. The preloaded transcript/genomic library adopted for the target search was the *E. grandis* JGI genomic project, *Phytozome*

v8.0, with internal number 201. The list of miRNAs predicted target genes obtained using the *psRNATarget* software is described in Supplementary Table S1.

2.5.1. Primer design of target genes

The primers used for target genes were designed using *Primer3 plus v2.3.3* software (Untergasser *et al.*, 2012; *http://primer3plus.com/cgi-bin/dev/primer3plus.cgi*) in order to flank the predicted miRNA binding sites. For each gene, it was necessary design two primers to 3´end, such as the first is a reverse gene-specific primer (GSP) and the second a reverse nested gene-specific primer (Nested GSP), being the latter characterized as more specific. The GSP primers were designed according to the following conditions: i) 50 - 70% GC content to obtain a high annealing temperature (> 72ºC), ii) 23 - 28 nucleotides in length to increase specificity of binding, iii) low GC content at 3′ends to minimize extension by DNA polymerase at non-target sites (no more than two G or C residues in the last five bases), iv) no self-complementary sequences within the primer or no sequence complementary to the primers supplied in the kit, especially at the 3' end, and v) the annealing temperature greater than 72ºC. For the Nested GSP design conditions were similar to those previously described, bearing in mind that only far enough from the original GSP so that you can distinguish the products of original and nested PCR by size. The primers sequences obtained as described in Supplementary Table S2.

2.6. Experimental validation of miRNAs target genes

2.6.1. mRNA isolation from total RNA

The total RNA was extracted from the reaction wood model (RW_{2010}) tissues, and the cleansed fraction of mRNA was obtained by purifying 1 µg of RW_{2010} (TW and OW) total RNA pool using the MicroPoly (A) Purist™ kit (Ambion®, Austin, TX, USA), following the manufacturer's instructions: water diluted total RNA was combined with the proprietary Binding Solution, and a pre-measured aliquot of Oligo (dT) Cellulose was added to it. The mixture was incubated with continual shaking, allowing hybridization between the poly (A) sequences found on most mRNAs and the Oligo (dT) Cellulose. The Oligo (dT) Cellulose was then transferred to a Spin Column and washed to remove nonspecifically bound material and ribosomal RNA. Finally, the poly (A) RNA was eluted using pre-warmed THE RNA Storage Solution.

2.6.2. RNA ligase-mediated 5´ rapid amplification of cDNA ends (RLM 5'-RACE)

For identification of products of the cleavage site within the miRNAs targets, a modified RNA ligase-mediated 5' rapid amplification of cDNA ends (RLM 5'-RACE) protocol was performed, using the GeneRacer™ Kit (Invitrogen®, Carlsbad, CA, USA). The Figure 8 showed the summarized flowchart of this procedure. The 5' RNA oligo adaptor (5' CGA CUG GAG CAC GAG GAC ACU GAC AUG GAC UGA AGG AGU AGA AA 3') was ligated to approximately 150 ng of mRNA using T4 RNA ligase (Invitrogen®, Carlsbad, CA, USA). The ligated mRNAs were then reverse transcribed using SuperScript™ III reverse transcriptase and the random primers to create RACE-ready first-strand cDNA with known priming site at the 5' end (Invitrogen®, Carlsbad, CA, USA).

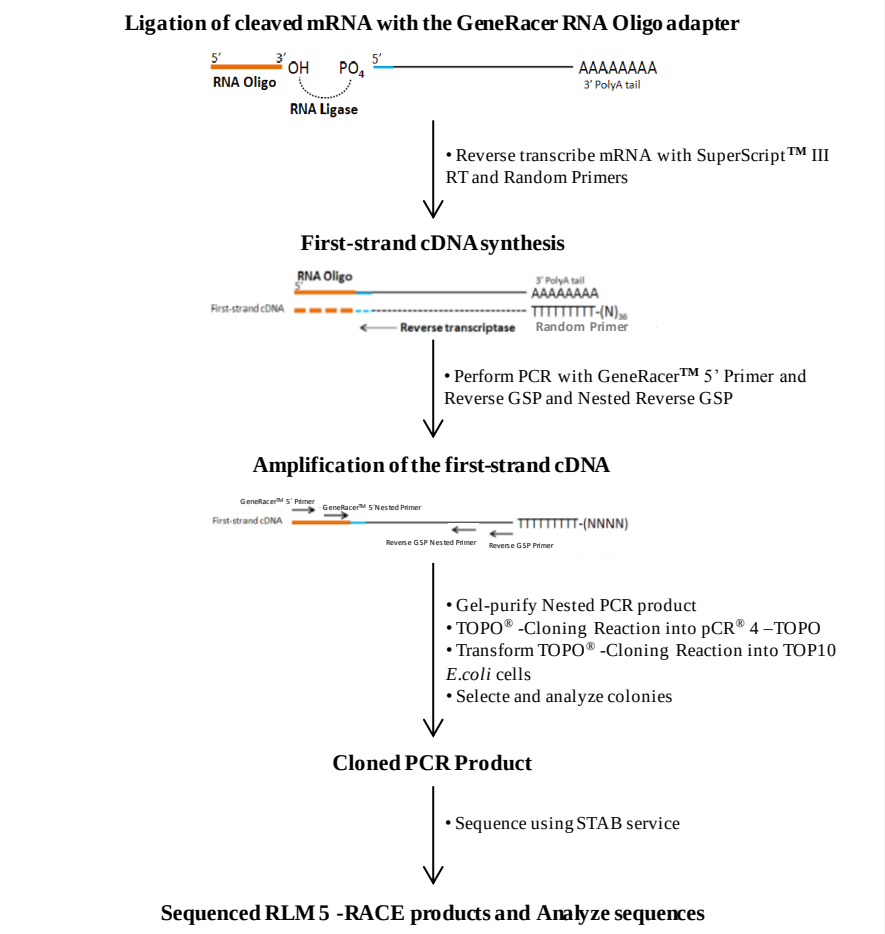


Figure 8: Flowchart summarizing the GeneRacer™ experimental outline. Initially, the GeneRacer™ RNA Oligo adapter was ligated to full-length mRNA. This mRNAs ligated were then reverse transcribed using SuperScriptTM III reverse transcriptase and the random primers to create RACE-ready first-strand cDNA. For obtain 5' ends, was amplified the first-strand cDNA using the GeneRacer™ 5´ primer and 5´nested PCR with respective reverse GSP primer or nested primer. The final nested PCR products with the expected size were cloned using the TOPO® TA Cloning® Kit (Promega®, Hollow Road; Madison, WI, USA). The transformed colonies were identified for colony PCR and for each PCR product a minimum of ten independent colonies were sequenced at STAB Vida (STAB Vida, Lisboa, Portugal), and afterwards analyzed with appropriate software (Source: GeneRacer™ Kit, Invitrogen® Carlsbad, CA, USA).

To obtain 5' ends, the first-strand cDNA was amplified using two rounds of RLM 5'-RACE reaction: the first PCR reaction was performed using GeneRacer™ 5' primer provided in the kit and one reverse gene-specific primers (reverse GSP) and the second PCR reaction was performed using the GeneRacer™ 5' nested primer and one reverse nested GSP (Figure 9). The PCR amplification program consisted of an initial denaturation cycle at 94ºC for 2 min, followed by 5 cycles with the following conditions: denaturation cycle at 94ºC for 30 s, annealing at 72ºC for 1 min, extension of primers at 60ºC for 30s and annealing at 70ºC for 1 min. This was followed by 35 cycles of 30s and 1 min with conditions as above. To conclude, we performed a final extension at 68ºC for 20 min. The evaluation of the PCR products was performed in 2% (w/v) agarose gel and stained with SYBR safe (Invitrogen®, Carlsbad, CA, USA). The molecular marker used was 1 Kb Plus DNA Ladder (Invitrogen®, Carlsbad, CA, USA), and the bands visualized in a *Gel Doc-1000 UV* image acquisition system (Bio-Rad® Laboratories, Inc.).

To confirm the pattern of miRNA-directed cleavage, the final PCR products with correct expected size (See Supplementary Table S2) were cloned using the TOPO® TA Cloning® Kit (Invitrogen®, Carlsbad, CA, USA). The transformed colonies were identified and handpicked for colony PCR. For each PCR product a minimum of ten independent colonies were sequenced at STAB Vida (STAB Vida, Lisboa, Portugal). The results were analyzed using the *BioEdit Sequence Alignment Editor* 7.0.5.3 system (Hall, 1999; Ibis Biosciences®, Carlsbad, CA, USA) to check if the 5´end of the sequenced product matched the predicted miRNA cleavage site.

2.7. Gene expression profiling of selected miRNAs target genes

Based on evidence indicated by RLM 5'-RACE validation, we decided to test the target genes for which the results indicated the correct expected cleavage site and for which demonstrated non predicted cleavage site in terms of gene expression profiling in a set of developing xylem tissues. A total of three target genes comprising of three miRNAs candidates families were selected (Table 7). These targets were subject to an extensive expression profile in a wider panel of developing xylem tissues by quantitative real-time PCR (RT-qPCR). *PikoReal Real-Time PCR* (Thermo Scientific™, Wilmington, Delaware, USA) was used for the simultaneous quantification of 96 individual RT-qPCR reactions, including the genes of interest, reference genes and negative control. The reference genes should present at the minimum possible variation over the samples analyzed. These genes are usually *housekeeping* genes, which integrate the set of genes underlying the expression of proteins that sustain the basic functions required for cell maintenance (Demidenko *et al*., 2011; Morgante *et al*., 2011). These genes should be constitutively transcribed at a relatively

constant level throughout various tissues and conditions (Nicot *et al*, 2005). Cassan-Wang *et al*. (2012) reported 21 as good reference genes for *E. globulus*. Based on this, we selected a total of five reference genes: i) four reference genes (PP2A1, PP2A3, PTB and SAND) for the SW model, and ii) four reference genes (Helicase, PP2A1, PP2A3 and PTB) for RW_{2011} model (Supplementary Table S3), using the *geNorm* (Vandesompele *et al*., 2002) and *NormFinder* (Anderson *et al*., 2004) software.

Table 7: Biological validation selected miRNAs function and selection criteria.

MiRNA Candidate	Gene Family	Selected Target Gene	Putative Function (*psRNATarget* server)	Selection Criteria
miR171	MIR171	Eucgr.E01509	GRAS family transcription Factor	Cleavage pattern confirmed
miRCa-04	MIR396	Eucgr.C01382	Unknown	Cleavage pattern confirmed
miRCa-08	MIR172	Eucgr.J02113	Related to AP2.7	Cleavage pattern confirmed

2.7.1. Plant material

Two sets of tissues from *E. globulus* species were used, featuring contrasting biological conditions in developing xylem formation using two models of study (Figure 9). The total RNA was extracted according to the protocol described in subsection §2.2.1., and the equimolar pools of total RNA were gathered according to the type of tissue and bending time of each sample, according following described: i) for the seasonal wood (SW) model, four pools were formed by developing xylem sampling date corresponding to each season of the year, and ii) for the reaction wood (RW_{2011}) induction model, four pools were formed by developing xylem type (tension and opposite wood), corresponding to one week of bending and three/four weeks of bending and respectively sampling time.

2.7.2. Quantitative real-time PCR assay (RT-qPCR)

Before cDNA synthesis, trace quantities of DNA were eliminated using Ambion's Turbo DNA-*free*™ kit (Ambion®, Austin, Texas, CA, USA), according to the manufacturer's procedure (Figure 9). One µg of the RNA template (developing xylem from SW and RW_{2011}) was diluted in 10 µL of nuclease-free water and reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems®, Foster City, CA, USA), following the manufacturer's procedure. Each reaction was performed for a total of three biological replicates (three genotypes) for each condition, and occurred in a final volume of 20 µL by adding 10X RT buffer, 10X RT Random Primers, 25X dNTP Mix and 50 U/µL MultiScibe™ Reverse Transcriptase, according to the manufacturer's instructions.

PCR primers are described in Supplementary Table S4 and were designed using the *QuantPrime* software (*http://www.quantprime.de/*), according to the following conditions: *E. grandis* (*Phytozome 6.0*) as a reference organism, the quantification protocol was SYBR Green real-time qPCR (accept splice variant hits) in which the primer pairs are optimized for real-time SYBR Green quantification (60-150 bp), with an annealing temperature of 60ºC, guanine-cytosine (GC) content 40-60% and the transcript splice variants will not be considered as unspecific hits when testing for specificity. The reference genes as described in Supplementary Table S3 were selected based on a pre-screen of Helicase, PP2A1, PP2A3, PTB, SAND and UBQ14 genes (Cassan-Wang *et al*., 2012) on different developing xylem conditions using *geNorm* (Vandesompele *et al*., 2002) and *NormFinder* (Anderson *et al*., 2004) by *GenEx* software (version 4.3.8) (MultiD, Göteborg, Sweden). After performing a RT-qPCR to determine the expression pattern and its stability as internal control, we selected four reference genes based on their best performance (less variation in the wide spectra of samples selected for this assay).

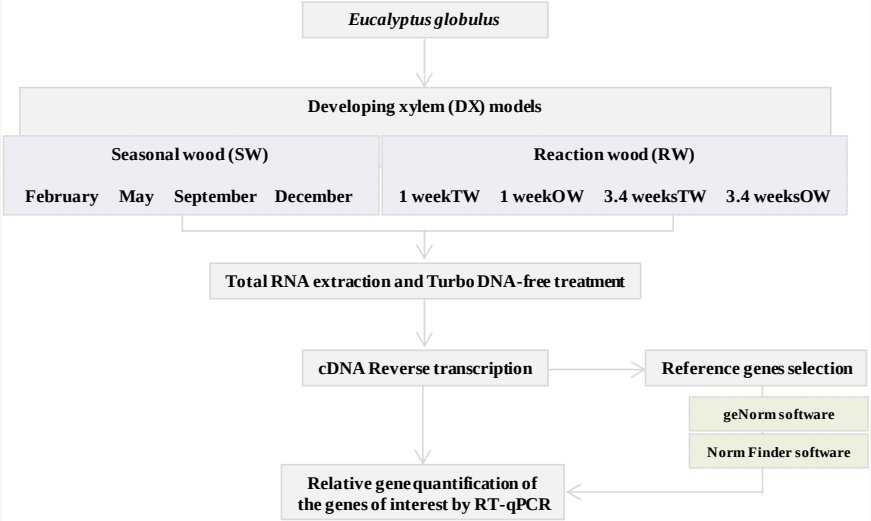


Figure 9: Summarized scheme describing the main steps of miRNA target genes expression profiling starting from total RNA extraction to quantification of gene expression on *E. globulus*.

The RT-qPCR reactions were performed in a *PikoReal Real-Time PCR* system (Thermo Scientific™, Wilmington, Delaware, USA), by adding 5 µL of SYBR Green Supermix (Bio-Rad® Laboratories, Inc.), 1 µL of diluted cDNA template (1:10), 0.2 µM of each primer, and nuclease-free water to a final volume of 10 µL. A negative control was run without cDNA template with every assay to assess the overall cross contamination and primer dimerization. After one initial incubation step at 95°C for 30s, amplifications were performed for 40 cycles with the following cycle profile: a denaturing step at 95°C for 15s followed by annealing at 60ºC for 15s and extension step at 72ºC for 30s. Fluorescence data were collected during

the 60ºC step, and the specificity of products was confirmed by performing a melting temperature analysis at temperatures ranging from 65°C to 95°C in intervals of 0.5°C. For all genes studied, three independent reverse transcription (RT) reactions were performed and pooled for each condiction. A total of three technical replicates were analyzed for each condition and for each gene in study.

2.7.3. Data analysis

Following raw data acquisition, we determined the amplification data (RFU) on background-subtracted mode. The specific efficiency of amplification was determined individually for each gene with *LinRegPCR* software (version 11.0, 2012). Amplification efficiency (E) expressed as a value between 1 and 2, and is calculated from slope of the amplification curve in the exponential phase. Ideally the PCR efficiency is 100%, meaning that in each cycle the amount of amplicons doubles. PCR efficiency per amplicon is assumed to be constant, but individual samples have slightly variable PCR efficiencies, so it was used the mean PCR efficiency per amplicon in the analysis. To allow this, the samples have to be assigned to amplicon groups, and the average efficiency of all samples for each amplicon group was used to do the quantification.

On the other hand, we determined the cycle threshold values (Ct), which translate the expression quantification of genes under study. This value is defined as the number of cycles required for values of the fluorescent signal to cross the threshold (i.e. exceeds background level). The Ct value of a reaction is determined by the amount of sample present at the start of the amplification reaction, where Ct levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. the lower the Ct level the greater the amount of target nucleic acid in the sample) (Pfaffl, 2001).

Using the *GenEx5* software (MultiD Analyses AB, Gothenburg, Sweden), C_t values obtained in the reading of each plate were processed and analyzed statistically as following:

1. Efficiency correction: from the efficiency value previously calculated for each of the genes, was possible to correct the specific efficiency, according to the equation $C_{tE=100\%}=C_{tE}\frac{\log(1+E)}{\log(2)}$, where C_{tE} is the uncorrected C_t value, and E is the efficiency.

2. Normalization of reference genes: from the expression of target genes (TG) could normalize them for the expression for the expression of reference genes (RG), according to the equation $C_{tnorm} = C_{tTGi} - \frac{1}{n}\sum_{i=1}^{n} Ct_{RGi}$.

3. Relative quantities: from the assignment of an amount of the maximum value of C_t, it was possible to calculate the relative expression of TG compared to the specified samples,

according to the equation $N_{Rel} = 2^{(Ctmin - Ct)}$. By selection the option *sample* works in the same way by comparing one given sample.

4. Descriptive statistics: from the three biological replicates in study, we can group them and get an average of three values for each TG in study, using descriptive statistics.

5. Student's t-test: for the amount on a specific sample, it was possible to assess whether there are significant differences of means and were selected exactly two samples groups to be compared on an interval t-test with 95% confidence ($P<0.05$).

3

Results

Contents

3.1. Nucleic acids extraction

3.1.1. Total RNA extraction, integrity and quantification

The Figure 10 displays the electrophoretic separation of RNA pools from developing xylem samples and other tissues and organs. The gel displayed sharp and clear RNA bands, presenting no visible signals of RNA degradation. It was possible to distinguish the two sharp rRNAs bands (28S and 18S) and small RNA bands were clearly visible in all samples. In the case of the leaves (see AdL and YoL) we found several other bands that probably correspond to other ribosome and plastidial RNAs.

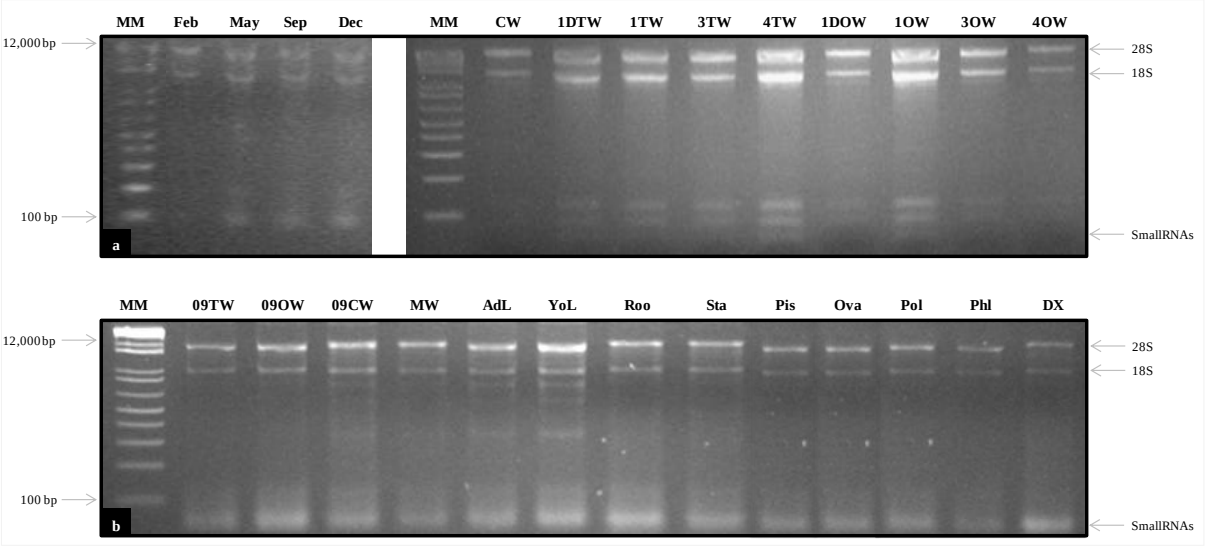


Figure 10: An example of the integrity of the total RNA extracted from *E. globulus* DX tissues. 2% (w/v) agarose gel (0.5X TBE) stained with *SYBR-safe* was used to test the quality of the total RNA. No visible signs of RNA degradation were observed, and it was possible to distinguish sharp of rRNAs bands (28S and 18S) attesting of RNA quality. (**a**) Total RNA samples from seasonal wood assay to formed the panels to Northern blot validation (left image), and the total RNA samples from reaction wood induction to formed RLM 5'-RACE libraries (right image). (**b**) Total RNA samples from reaction wood induction assay and various tissues/organs and developing xylem to formed the panels to Northern Blot validation. MM – Molecular marker 1 Kb Plus DNA Ladder (range from 12 kb to 100bp), and the pools codes was described in Table 4 and Table 6 (described in material and methods, subsections §2.3 and §2.4).

Total RNA spectrophotometric quantifications, using the *NanoDrop ND-1000* are shown in Table 8. In average we obtained 21.10 µg RNA/100 mg of fresh material, but the yield was highly variable, presenting the ovaries samples the minimum value (6.10 µg RNA/100 mg) and the tension wood (09TW) samples the maximum value (37.61 µg RNA/100 mg). For the set of seasonal wood (SW) samples, it was observed that there was a decrease of total RNA yield between February (15.79µg/100mg) and September (10.77µg/100mg) samples, and then a slight increase in December (13.08µg/100mg) sample. For the set of reaction wood (RW_{2009} and RW_{2010}) samples, it was presented a similar trend between the two types of tissues collected, showing a higher average of total RNA yield in the tension wood

(37.61µg/100mg and 27.54µg/100mg) than in opposite wood (36.48µg/100mg and 26.21µg/100mg), respectively to RW_{2009} and RW_{2010}. However, lower yield was observed for both controls (24.27µg/100mg and 18.71µg/100mg, respectively) and for mature wood (15.53µg/100mg) samples, comparatively to tension and opposite wood samples. Finally, the set of tissues and organs samples presented a high variation of total RNA yield among the samples. The lowest RNA yield was obtained for the ovaries (Ova) sample (6.10µg/100mg) and the highest yield was obtained for the adult leaves (AdL) sample (47.31µg/100mg).

Table 8: Quantification of total RNA samples extracted from selected *E. globulus* tissues, using a *NanoDrop ND-1000* spectrophotometer.

Plant Material		**RNA Yield (µg/100mg plant material)**	**Ratios (nm)**	
			$A_{260/280}$	**$A_{260/230}$**
Seasonal wood (SW)	Feb	15.79	2.06	2.04
	May	14.14	2.05	2.04
	Sep	10.77	2.04	2.04
	Dec	13.08	2.04	2.02
Reaction wood (RW_{2009})	09TW	37.61	2.10	2.02
	09OW	36.48	2.09	2.01
	09CW	24.27	2.08	2.00
Reaction wood (RW_{2010})	TW	27.54	2.02	2.00
	OW	26.21	2.06	2.01
	CW	18.71	2.05	2.01
Mature wood (MW)	MW	15.53	2.07	2.00
Tissues and organs	AdL	47.31	2.07	2.05
	DX*	20.69	2.05	2.01
	YoI	28.36	2.09	2.03
	Sta	16.65	2.10	2.01
	Ova	6.100	1.98	1.10
	Pol	27.48	2.03	2.00
	Pis	13.46	2.06	1.14
	Roo	13.24	1.83	1.17
	Phl	12.46	2.12	1.22

Legend: Feb – February: Sep – September; Dec – December; TW – Tension wood; OW – Opposite wood; CW – Control wood; AdL – Adult leaves; DX - Developing xylem; YoL – Young leaves; Sta – Stamens; Ova – Ovaries; Pol – Pollen; Pis – Pistils; Roo – Roots; Phl – Phloem.

The $A_{260/280}$ ratio tends to confirm the purity of all samples as most values presented a ratio equal or above a cut-off value (2.00) measured. However, ovaries and roots samples are an exception showing ratios of 1.98 and 1.83, respectively, but yet very acceptable. The majority of samples showed $A_{260/230}$ ratio equal or greater to 2.00, indicating low contamination with polysaccharides. Nevertheless, the ratio found for the ovaries (1.10), pistils (1.14), roots (1.17) and phloem (1.22) samples, reflecting the possible excess of polysaccharides of these samples. This particular contamination may have negative implications in the electrophoretic mobility of RNA samples. The Figure 11 exemplifies the results obtained after the staining and running 15% (w/v) polyacrylamide gel used to make the Northern blot membranes, for tissues and organs filter. In this figure an electrophoretic distortion is visible for pistil's samples in agreement with the previous observation of low $A_{260/230}$ ratio (see Table 8).

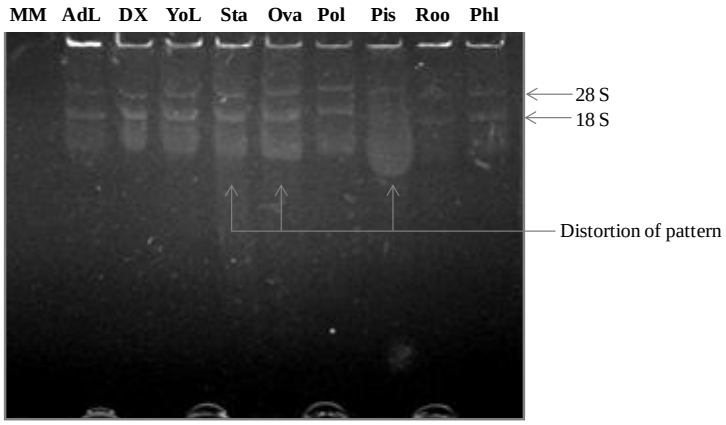


Figure 11: An example of the total RNA samples extracted from *E. globulus* DX tissues/organs. Electrophoresis 15% (w/v) polyacrylamide gel (1X MOPS), stained with SYBR-Safe was used for further analysis by Northern Blot. It is possible to distinguish rRNAs bands (28S and 18S) and the distortion of electrophoretic pattern. MM – Molecular marker synthetic RNA of 19 nt and 24 nt. The pools codes is described in Table 6 (see Material and Methods, subsection §2.4.1).

3.2. Northern blot validation of selected miRNAs candidates

All eight miRNAs candidates were analyzed by Northern blots hybridizations since in each, a band with the expected size (approximately 20-24 nt) was observed (Figures 12 and 13). Because members within each miRNA family share high sequence homology, the hybridization signals could reflect the transcript levels of the family rather than a single member. The increasing of the bands size observed in some gels pictures, such as miRCa-02 and miRCa-04 forms (Figure 12), could be associated with electrophoresis distortion. However, the size of bands in both miRCa-04 and miRCa-08 forms (Figure 13), a second band was observed in young leaves (YoL) and developing xylem (DX) samples with about 25-27 nt. This extra band may correspond to the so-called long-miRNAs previously described in *Arabidopsis* by Vazquez *et al*. (2008). According to these authors, long-miRNAs are processed by DCL3 from the same miRNA precursor used by DCL1 to generate 21 nt miRNAs, and has been shown to be regulated under certain conditions. Several long-miRNAs were conserved in different Angiosperm families throughout evolution, which suggests that it may also have a biological role in plants (Vazquez *et al*. 2008).

The observed intensity for each miRNA band was subsequently measured and normalized using U6 signal to create the histograms shown in Figure 12 and Figure 13. Distinct expression profiles were revealed, exposing different patterns of abundance variation for miRNAs tested within the selected tissue panels, suggesting a biological differentiation impact. Some candidates such as miRCa-09 forms presented low signal intensity in all assayed tissues, while others such as miRCa-04 and miRCa-08 presented consistent, exuberant transcript abundance in most assayed tissues. The miRCa-02 revealed the biggest differences between the transcript abundance obtained for both forms tested.

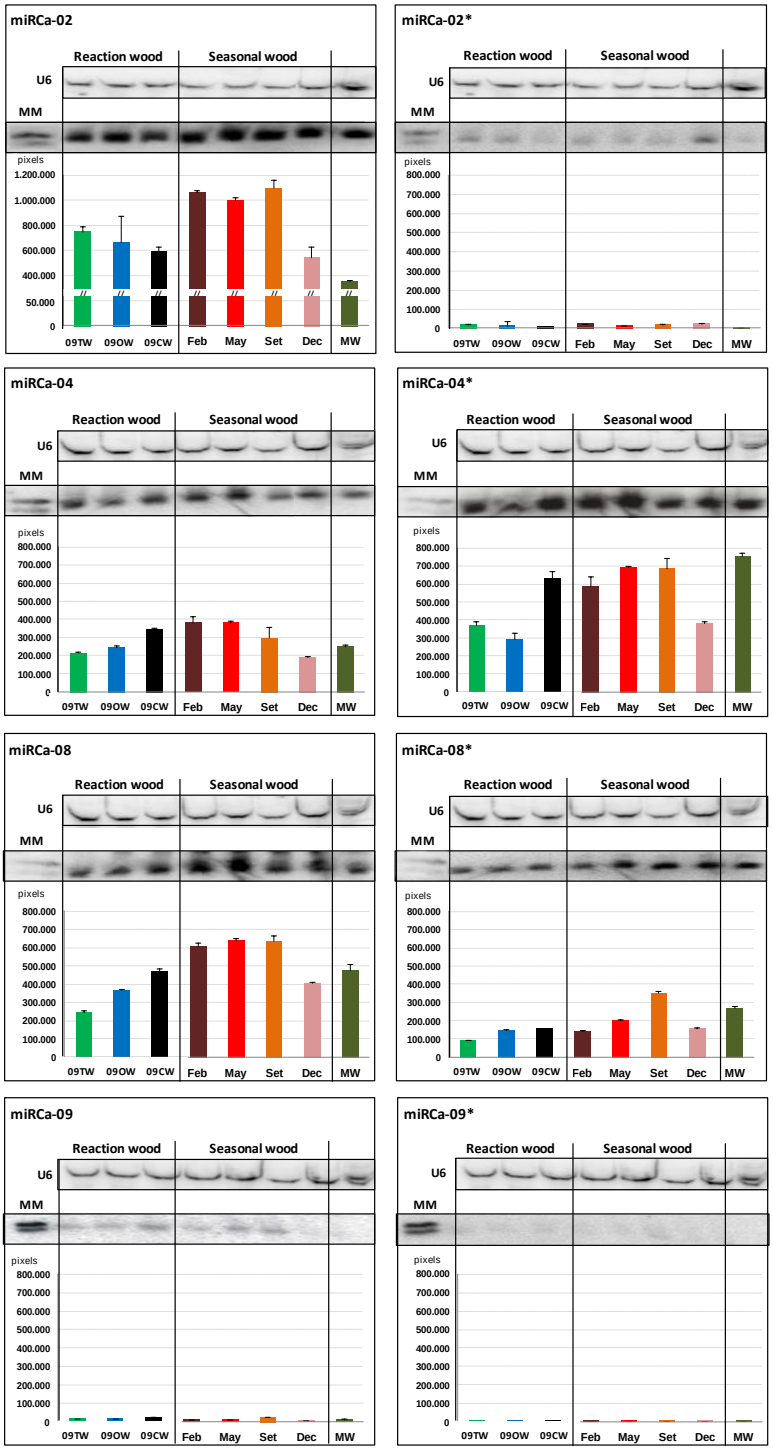


Figure 12: Expression pattern of four miRNAs candidates in developing xylem tissue panel (DX-panel), harvested under the *E. globulus* seasonal wood (SW) and reaction wood (RW_{2009}) models. The miRCa-02 form showed a high expression level in all tissues in study, but in miRCa-02* form presented a very low or residual expression variation. The miRCa-04 form showed a similar abundance and intensity in all tissues. The miRCa-04* form was more intense, and was preferentially expressed in control wood (09CW), February (Feb), May and September (Sep), but also showed a high expression level in mature wood (MW). The miRCa-08 form showed an increased expression in reaction wood model and a decrease in December (Dec) in comparison with other seasonal wood samples. The miRCa-08* form presented low expression in all tissues, highlighting an increase in September. Finally, miRCa-09 presented a very low expression level, being barely detectable in miRCa-09* form. MM – microRNA size marker from New England Biolabs (the darker bands correspond to 24 and 21 nt). Small nuclear RNA U6 was used as a loading control (upper bands in the pictures). The accumulation of all miRNAs was quantified using the *ImageJ 1.45s* software and normalized to the loading control. Expression values were calculated in comparison to U6 loading control.

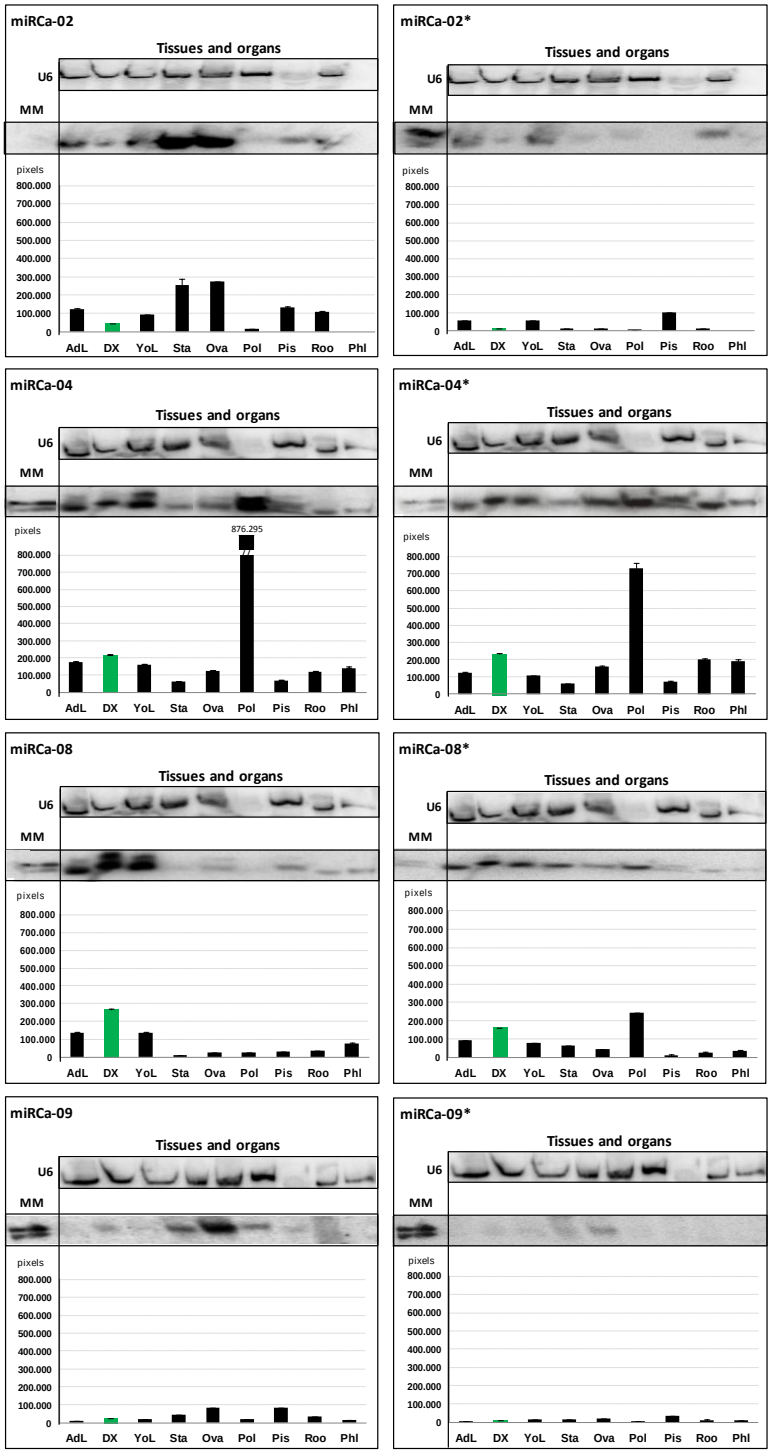


Figure 13: Expression pattern of four miRNAs *E. globulus* candidates in other tissues and organs panel (T/O-panel). The expression of miRCa-02 form was highly in stamens (Sta), ovaries (Ova), pistils (Pis), roots (Roo) and adult leaves (AdL), but also showed a lower expression in pollen (Pol) and developing xylem (DX) tissues. On the other hand, miRCa-02* showed a low expression variation, with the intensity of bands down, standing out a bit only in adult (AdL) and young (YoL) leaves and pistils. In both miRCa-04 forms, the expression stands in pollen, being lower and similar in other tissues and organs in study. The miRCa-08 form was highly expressed in DX, AdL and YoL. The miRCa-08* form showed preferentially expression in Pol and DX tissues, but also showed an expression variation in AdL, YoL and Sta tissues. The miRCa-09 form showed a lower expression level, being only slightly higher in Ova and Pis. In miRCa-09* forms showed a very residual expression. MM – microRNA size marker from New England Biolabs (the darker bands correspond to 24 and 21 nt). Small nuclear RNA U6 was used as a loading control (upper bands in the pictures). The accumulation of all miRNAs was quantified using the *ImageJ 1.45s* software and normalized to the loading control. Expression values were calculated in comparison to U6 loading control.

In the developing xylem panel (DX-panel, Figure 12), the strong signal intensities revealed in several developing xylem tissues for miRCa-04 and miRCa-08 suggests their functional association with vascular cambium differentiation activities and therefore an active regulatory role in *E. globulus* xylogenesis. Considering the reaction wood (RW) model, remarkable levels of variation were detected for miRCa-02, miRCa-04 and miRCa-08 suggesting their involvement in the biological response to gravitropic stimulation. The exception to this abundance gradient was verified for miRCa-04* form presenting higher miRNA abundance in control wood (09CW) tissue comparatively to tension wood (09TW) and opposite wood (09OW) tissues. Concerning the seasonal wood (SW) variation, all four miRNAs candidates have shown consistently their lowest peak of abundance in developing xylem tissue collected in December. In most cases, the miRNAs abundance levels were similar in February and May harvested points, and slightly lower in December.

In the other tissues and organs panel (T/O-panel, Figure 13), the guide and star* forms often differed in their patterns of transcript abundance. The exception was miRCa-04 since both forms revealed a similar profile, featuring an overabundance in pollen and in developing xylem tissues. In all miRNAs, the four floral tissues exhibited preferential abundance, at least one of these tissues presented a candidate overabundance comparatively to the remaining tissues: stamens (miRCa-02 form), pistils (miRCa-02* form and both miRCa-09 forms), ovaries (miRCa-02 and miRCa-09 forms), and pollen (both miRCa-04 forms and miRCa-08* form). Some miRNAs candidates revealed important or even preferential presence in other tissues such as roots (miRCa-02 form), developing xylem (both miRCa-04 forms and miRCa-08 form), adult and young leaves (miRCa-02 form) and phloem (both miRCa-04 forms). Only the miRCa-08 form revealed a preferential tendency of expression in developing xylem tissues. Most important, all miRNA candidates revealed differential presence in developing xylem tissues generated under the two xylogenesis models. The high presence of miRCa-04 and miRCa-08 forms in developing xylem tissues suggests that they could have an active important regulatory role in xylogenesis related processes.

3.3. Target gene prediction of selected miRNAs

To understand miRNAs potential functions, we used the selected four miRNAs candidates for guide form (miRCa-02, miRCa-04, miRCa-08 and miRCa-09). Additionally, the target gene prediction was also performed for miR171, selected to test the feasibility of target validation methodology (see material and methods, subsection §2.6.). The target gene predictions over the guide form of these five miRNAs candidates allowed us to reveal a total of 68 predicted miRNAs-cleaved transcripts encoded by 42 genes (Supplementary Table S1): two predicted target genes for miR171, four predicted target genes for miRCa-02, seven

predicted target genes for miRCa-04, twelve predicted target genes for miRCa-08, and finally seventeen predicted target genes for miRCa-09. Concerning these predicted targets, 48% of these target genes were found to integrate transcription factor gene families either with known or predicted functions, and 43% were associated with metabolic and cellular processes and also stress/defense functions. Additionally, 9% of the predicted target genes are annotated with unknown function, suggesting possible new roles for conserved miRNAs in this tree species (detail in Supplementary Table S1). Interestingly, miRNA-mediated target cleavage was found to be the prevalent regulation mechanism (reaching 97% of the regulation events predicted for miRNA candidates) over miRNA-mediated translation repression (3%). Predictions also suggested that each miRNA mediated the regulation of a multiplicity of genes as each miRNA candidate predicted to target more than one gene. The most extreme situation was found for candidate miRCa-08 predicted to target 25 transcripts encoded by 12 genes, and for candidate miRCa-09 that it was predicted to target 22 transcripts encoded by 17 genes. In opposition, predicted target genes for some of the conserved miRNAs candidates were found to be homologs of genes previously identified in other plant species: GRAS transcription factor for miR171 (Eucgr.E01509), or APETALA2 (AP2) for miRCa-08 (MIR172; Eucgr.J02113) (Supplementary Table S1).

3.4. Experimental validation of miRNAs target genes

Seeking confirmation for our *in silico* cleavage pattern predictions, we have performed a RNA ligase-mediated 5' rapid amplification of cDNA ends (RLM 5'-RACE) protocol, as described in material and methods, subsection §2.6.2. This technique has been used to validate miRNAs targets as it allows us to detect the miRNAs cleavage site that generally occurs between the tenth and the eleventh nucleotides from the 5' end of the miRNA (Llave *et al*., 2002a; Kasschau *et al*., 2003). Considering the results generated in target genes prediction for this experimental validation, we selected three target genes for each miRNAs candidates. The selection criteria were to pick target gene presenting the lowest expectation value (higher confidence) or to pick target gene associated to an interesting putative function within the context of xylogenesis related processes. For the miR171 candidate, all target genes predicted were selected. The primers corresponding to all selected target genes can be found in Supplementary Table S2.

We started the validation work for all miRNAs candidates using the TW_{2010} library, and then proceeded to use the OW_{2010} library in cases in which no target cleavage was obtained. No PCR amplification products were obtained for the four target genes (Eucgr.H01257, Eucgr.D01126, Eucgr.G02793 and Eucgr.J00040). A single PCR amplification product was obtained for six target genes (Eucgr.E01509, Eucgr.K00370, Eucgr.B01226, Eucgr.B02279,

Eucgr.J02113 and Eucgr.H03865) and, finally for four target genes (Eucgr.Eucgr.E01875, Eucgr.C01382, Eucgr.A01182 and Eucgr.C02716) two fragments were obtained (Table 9). Considering these results, experimental validation proceeded for five target genes showing expected size amplification (see targets marked in the Table 9 with "plus" symbol).

Table 9: List of fourteen selected target genes with the information about PCR reaction of 5´nested products and the respective expected size of the bands.

			PCR 5´Nested Products			Cleavage site pattern	
miRNA name	**Target gene**	**PCR products**	**PCR product observed size (bp)**	**PCR product expected size**	**Cloning and Sequencing**	**Predicted**	**Non Predicted**
miR171	Eucgr.E01509	1	324	**+**	**Yes**	**Validated**	No
	Eucgr.H01257	0	176	-	No	No	No
miRCa-02	Eucgr.E01875	2	269/300	**+ / -**	**Yes**	No	**Yes**
	Eucgr.K00370	1	101	-	No	No	No
	Eucgr.B01226	1	93	-	No	No	No
miRCa-04	Eucgr.B02279	1	218	**+**	**Yes**	No	No
	Eucgr.C01382	2	329/450	**+** / -	**Yes**	**Validated**	**Yes**
	Eucgr.D01126	0	171	-	No	No	No
miRCa-08	Eucgr.J02113	1	86	**+**	**Yes**	**Validated**	No
	Eucgr.A01182	2	100/179	-	No	No	No
	Eucgr. G02793	0	68	-	No	No	No
miRCa-09	Eucgr.H03865	1	101	-	No	No	No
	Eucgr.C02716	2	100/200	-	No	No	No
	Eucgr.J00040	0	501	-	No	No	No

Legend: (-) expected and observed sizes (bp) were different; (+) expected and observed sizes (bp) were equals.

Validation of the miR171 targets (Eucgr.E01509 and Eucgr.H01257 target genes):

The miR171 target validation was performed over two putative target genes (Eucgr.E01509 and Eucgr.H01257), using the TW_{2010} library. Only the PCR amplification of the target gene Eucgr.E01509 (GRAS transcription factor) yielded a fragment with the expected size (approximately 324 bp), which was later cloned and sequenced. The sequences alignment showed that the single cloned fragment aligned correctly confirming the expected cleavage site transcript position (approximately 1686 bp), thus validating Eucgr.E01509 as a miR171 target gene (Figure 14). In opposition, since no PCR product was obtained for Eucgr.H01257 (GRAS transcription factor), the predicted post-transcriptional regulation of this putative target gene, mediated this miRNA remains without confirmation.

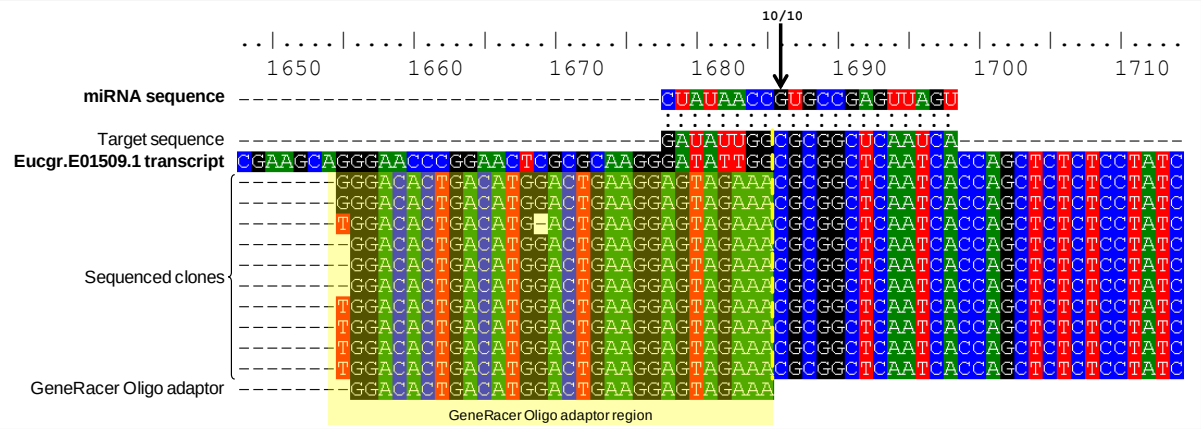


Figure 14: Validation of *in silico* predicted Eucgr.E01509 target gene for miR171 candidate by RLM 5'-RACE. The predicted cleavage site was validated after sequencing a fragment with 324 bp (correct expected size), using the TW_{2010} library (DX tissue). The upper lines represent the miRNA sequence (3' – 5'), the target sequence (5' – 3') and the respective *E. grandis* transcript. The position of the miRNA sequence and its complementary target sequence are indicated. Vertical dashes indicate a Watson-Crick base-pair. The arrows above the miRNA sequence indicate the 5' ends of cleavage products and the number of independent clones that mapped the site of cleavage to that position.

Validation of the miRCa-02 targets (Eucgr.E01875, Eucgr.K00370 and Eucgr.B01226):

Regarding miRCa-02 candidate, the validation of three putative target genes (Eucgr.E01875, Eucgr.K00370 and Eucgr.B01226) was assayed using TW_{2010} and OW_{2010} libraries. For the target gene Eucgr.E01875 (Unknown function) two fragments were amplified: one presenting approximately 269 bp (matching the expected size) using the TW_{2010} library and the other presenting approximately 300 bp using the OW_{2010} library. Both bands were cloned and sequenced. The sequencing results even thou it allowed us to confirm the correct transcript identity for both fragments, it provided no confirmation of the predicted pattern of transcript cleavage for miRCa-02 candidate (approximately the 177 bp transcript position). The two sequenced cloned products each grounded in several cloned copies suggests the existence of two non predicted cleavage sites: one in the TW_{2010} library amplified fragment featuring a cleavage site located at 290 bp transcript position (Figure 15a), and other in the OW_{2010} library amplified fragment featuring a cleavage site located at 126 bp transcript position (Figure 15b). The sequencing of the remaining miRCa-02 target genes (Eucgr.K00370 and Eucgr.B01226) was discarded since none of the PCR products presented the expected size (101 bp and 93 bp, respectively).

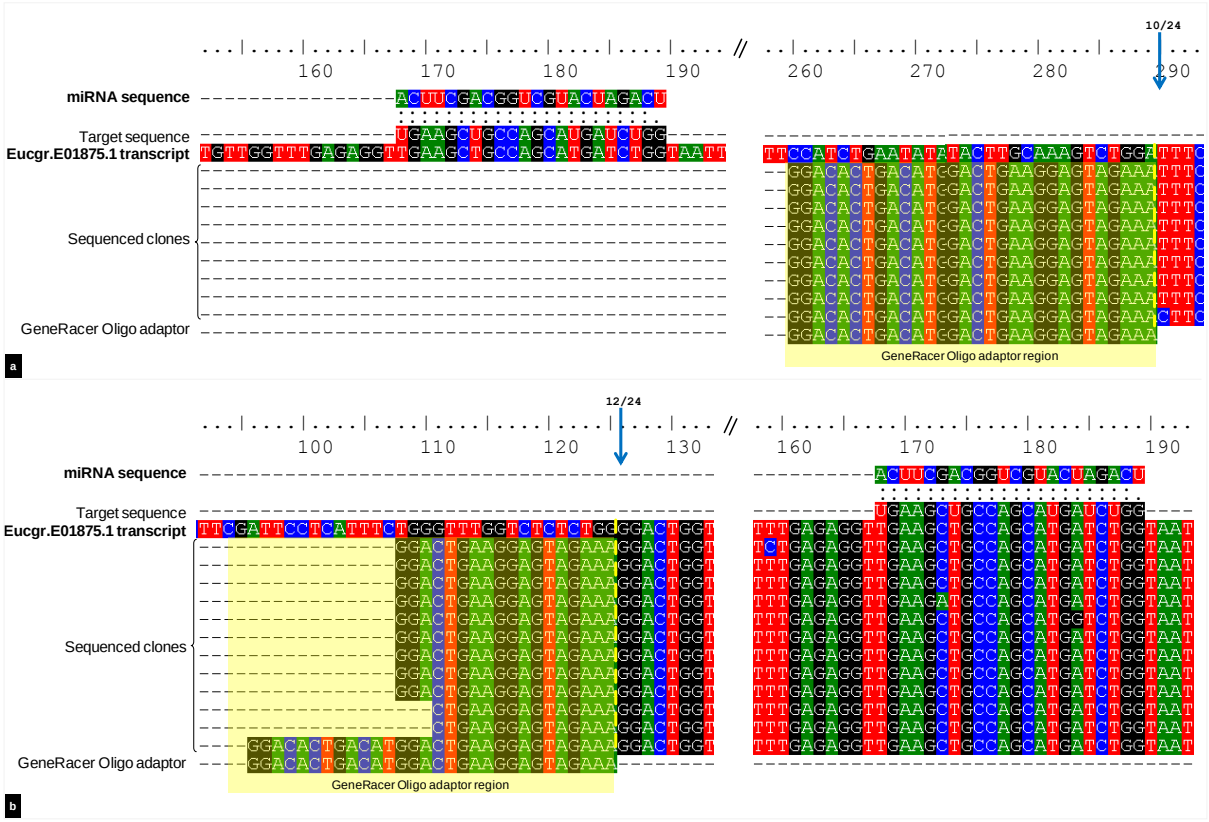


Figure 15: Validation of *in silico* predicted Eucgr.E01875 target gene for miRCa-02 candidate by RLM 5'-RACE. Two non predicted cleavage sites were revealed after sequencing (**a**) a fragment with 269 bp (correct expected size) using the TW_{2010} library (DX tissue), and (**b**) a fragment with 300 bp, using the OW_{2010} library (DX tissue). The upper lines represent the miRNA sequence (3' – 5'), the target sequence (5' – 3') and the respective *E. grandis* transcript. The position of the miRNA sequence and its complementary target sequence are indicated. Vertical dashes indicate a Watson-Crick base-pair. The arrows above the miRNA sequence indicate the 5' ends of cleavage products and the number of independent clones that mapped the site of cleavage to that position.

Validation of the miRCa-04 targets (Eucgr.B02279, Eucgr.C01382 and Eucgr.D01126):

For miRCa-04, three putative target genes: Eucgr.B02279 (Antiviral helicase SKI2), Eucgr.C01382 (Unknown function) and Eucgr.D01126 (Integral to membrane) were assayed using both TW_{2010} and OW_{2010} libraries. Only for Eucgr.B02279 and Eucgr.C01382 a fragment matching the expected size was obtained (218 bp and 329 bp, respectively). The cloning and sequencing of these fragments revealed two distinct situations. In the case of Eucgr.B02279, the sequencing of the fragment amplified using TW_{2010} library revealed the inexistence of the expected transcript homology and therefore was discarded from further analysis. In contrast, the sequencing of the Eucgr.C01382 fragment amplified using OW_{2010} library, provided the confirmation of the correct transcript homology and validated the predicted cleavage site at 99 bp transcript positions. Thus, this gene was validated as miRCa-04 target gene (Figure 16a). However, for this target gene non predicted cleavage

site at 110 bp transcript position has also been revealed. This result is very interesting since this non predicted cleavage site is situated nearby the predicted cleavage site (Figure 16b).

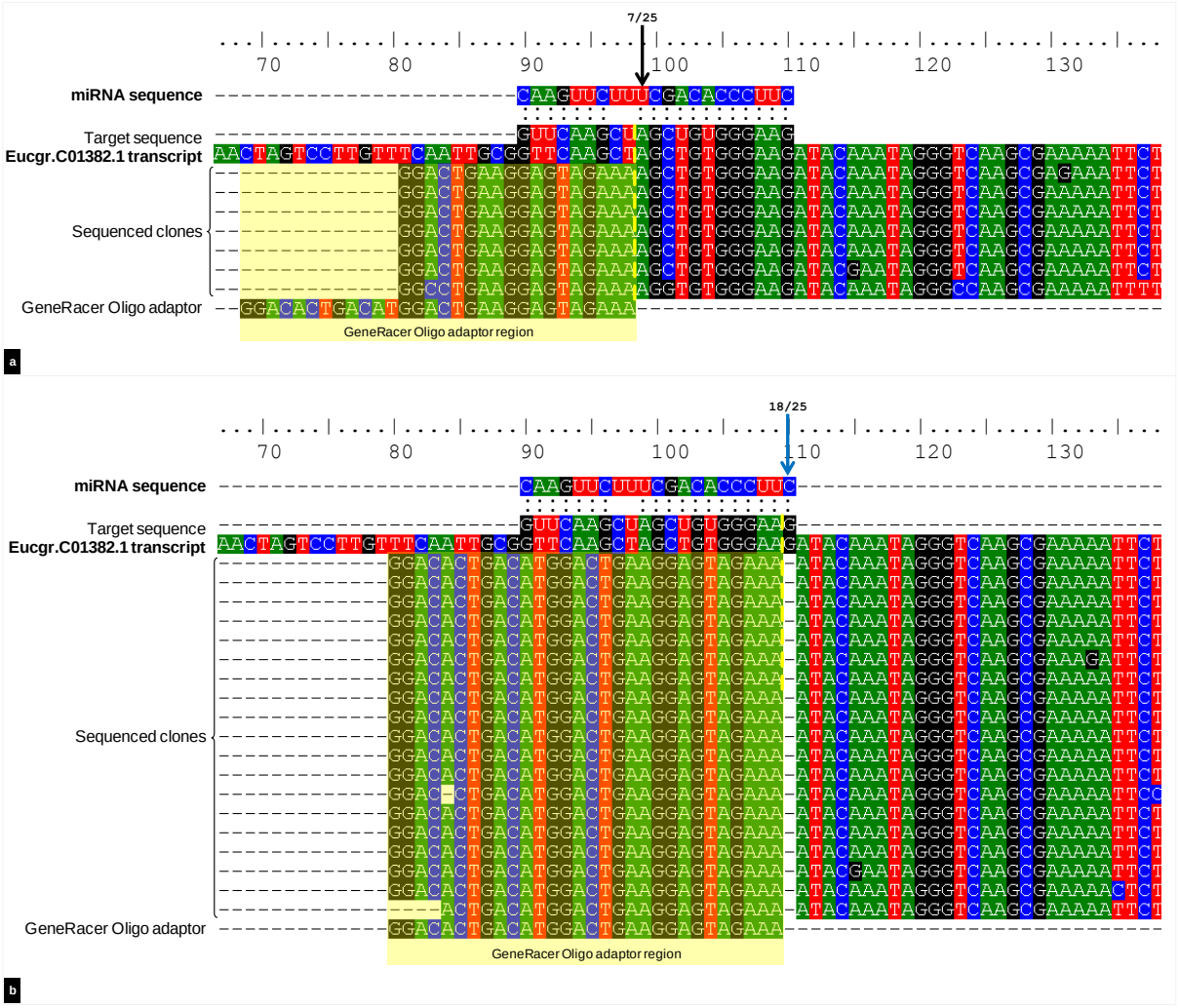


Figure 16: Validation of *in silico* predicted Eucgr.C01382 target gene for miRCa-04 candidate by RLM 5´-RACE. The predicted cleavage site was validated after sequencing (**a**) a fragment with 329 bp (correct expected size), using the OW_{2010} library (DX tissue). On the other hand, non predicted cleavage site revealed after the sequencing (**b**) a fragment with 218 bp using the OW_{2010} library (DX tissue). The upper lines represent the miRNA sequence (3' – 5'), the target sequence (5' – 3') and the respective *E. grandis* transcript. The position of the miRNA sequence and its complementary target sequence are indicated. Vertical dashes indicate a Watson-Crick base-pair. The arrows above the miRNA sequence indicate the 5' ends of cleavage products and the number of independent clones that mapped the site of cleavage to that position.

Validation of the miRCa-08 targets (Eucgr.J02113, Eucgr.A01182 and Eucgr.G02793):

For the miRCa-08, the validation work was conducted for three putative target genes: Eucgr.J02113, Eucgr.A01182 and Eucgr.G02793 (all annotated as Related to AP2.7), using the TW_{2010} library. Only for target gene Eucgr.J02113 a fragment with the expected size (86 bp) was amplified, cloned and sequenced. The sequence alignment results allowed us to

attest the correct homology and allowed us to validate the predicted cleavage site at 1780 bp transcript position. Thus, Eucgr.J02113 was validated as a miRCa-08 target (Figure 17).

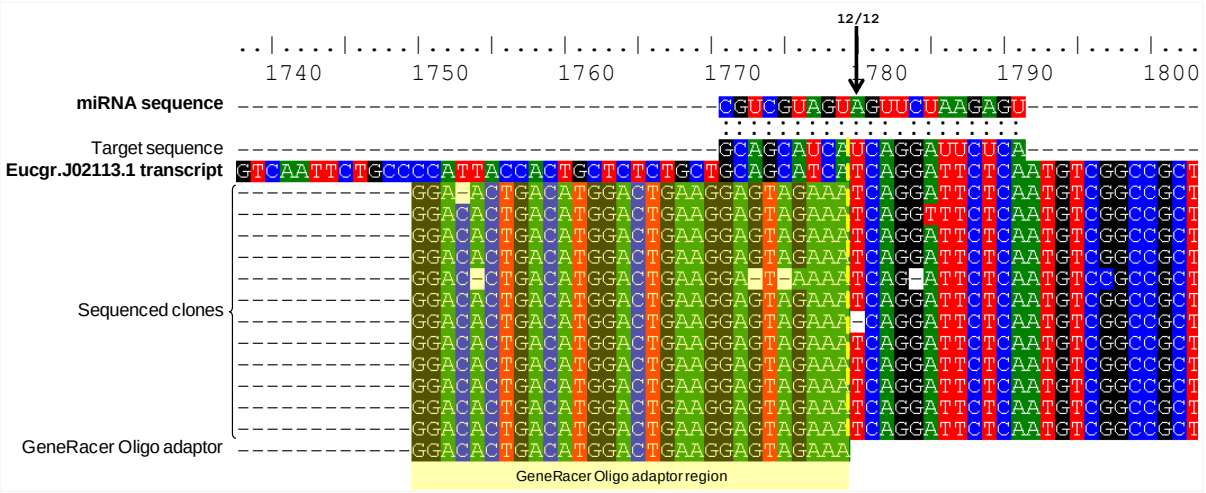


Figure 17: Validation of *in silico* predicted Eucgr.J02113 target gene for miRCa-08 candidate by RLM 5'-RACE. The predicted cleavage site was validated after sequencing a fragment with 86 bp (correct expected size), using TW_{2010} library (DX tissue). The upper lines represent the miRNA sequence (3' – 5'), the target sequence (5' – 3') and the respective *E. grandis* transcript. The position of the miRNA sequence and its complementary target sequence are indicated. Vertical dashes indicate a Watson-Crick base-pair. The arrows above the miRNA sequence indicate the 5' ends of cleavage products and the number of independent clones that mapped the site of cleavage to that position.

3.5. Gene expression profiling of selected miRNAs target genes

In order to evaluate the expression of miRNAs targets in developing xylem tissues generated under the two different *E. globulus* xylogenesis models, we profiled the expression of three conserved miRNAs target genes using RT-qPCR on a tissue panel which includes developing xylem samples from both xylogenesis models. The melting curve and the melt peak charts were calculated for all RT-qPCR reactions. Through the observation of a melt peak chart, we verified that a single perfect peak per reaction was obtained for Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113, suggesting that selective PCR amplification was achieved. The Supplementary Figure G1 shows the results obtained for amplifications. All curves have shown that the detection of the products occurred at approximately 77ºC (melting temperature), and also the absence of secondary PCR products contamination. The PCR efficiency was individually calculated for each primers pair using the C_t values obtained by RT-qPCR and *LinRegPCR* software (version 2012). The estimated PCR efficiencies are presented in Table 10. We considered as hallmark values, PCR efficiency values between 80% and 100%. Applying this rejection criterion, we noticed that all the PCR efficiencies revealed that these genes lie within this interval and therefore all the results are usable for this relative gene expression analysis.

Table 10: PCR efficiency determined for each assayed selected target gene and for all reference genes. Ct values were obtained by PikoReal Real-Time PCR system and analyzed by *LinReg v11.0* software.

Primers		Efficiency (%)	
		Seasonal wood (SW)	Reaction wood (RW_{2011})
Genes of Interest	Eucgr.E01509	87	86
	Eucgr.C01382	82	81
	Eucgr.J02113	84	82
Reference genes	Helicase	-	86
	PP2A1	86	89
	PP2A3	90	89
	PTB	83	80
	SAND	83	-

The expression pattern of the target genes was revealed for each model of study (SW and RW_{2011}). The Figures 19 and 20 display the results obtained for these analyses, and the Supplementary Table S5 details the results obtained by student's t-test statistical analysis. The RT-qPCR results demonstrated that Eucgr.E01509 (miR171 candidate), Eucgr.C01382 (miRCa-04 candidate) and Eucgr.J02113 (miRCa-08 candidate) target genes presented variations in both models of study, and that all these genes revealed expression levels throughout the assayed tissues (Figure 18 and Figure 19). However, the differences between expression levels were found not statistically significant ($P > 0.05$) in all tested tissues.

All genes presented a similar profile in seasonal wood (SW) model samples, namely a tendency to gene expression decrease from February to September, followed by an increase in December (Figure 18). February was always the month with higher gene expression.

Expression of the Eucgr.E01509 target gene (miR171 candidate):

This gene showed no significant expression differences ($P = 0.3117$) in the transition from February to May. However, between May and September a very significant decrease of expression ($P = 0.0017$) was detected as well as a statistically significant increase ($P = 0.0030$) for September and December.

Expression of the Eucgr.C01382 target gene (miRCa-04 candidate):

This gene revealed an expression decrease between February and May ($P = 0.0486$) and a marked decrease statically significant between May and September ($P = 0.0024$). On the other hand, between September and December a statistically highly significant ($P = 0.0002$) has also been detected.

Expression of the Eucgr.J02113 target gene (miRCa-08 candidate):

The expression of this target gene revealed no statistically significant differences (P = 0.9374) between February and May. However, between May and September a statistically significant decrease in gene expression has been revealed (P = 0.0014), as well as a significant increase in gene expression between September and December (P = 0.0069) was revealed.

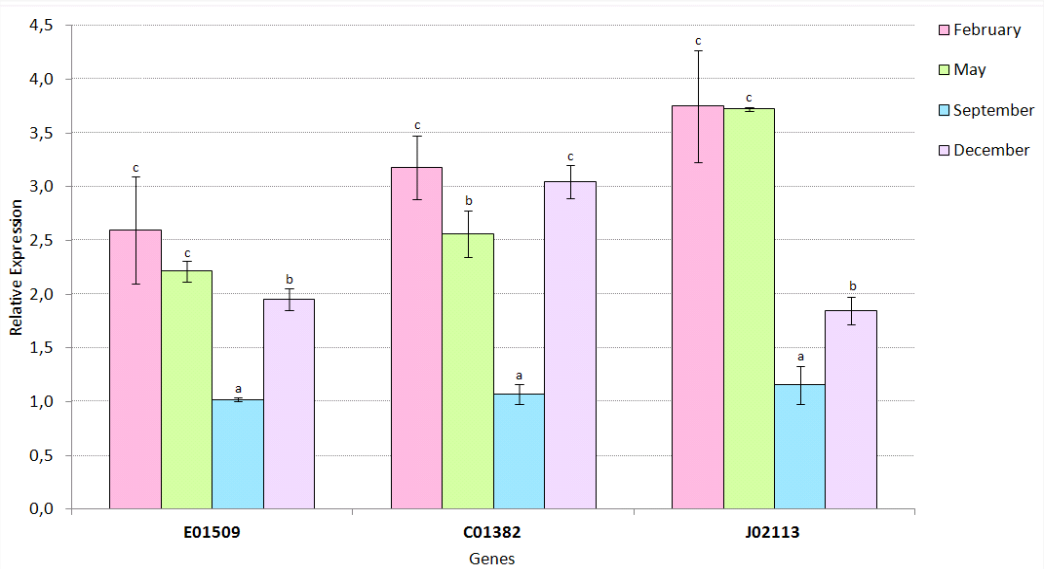


Figure 18: Quantitative real-time PCR (RT-qPCR) expression profiling of three *E. globulus* miRNAs target genes predicted and previously validated by RLM 5'-RACE methodology. February (pink), May (green), September (blue) and December (purple) are samples of developing xylem tissue, provided from seasonal wood (SW) model, respectively. Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113 are target genes of miR171, miRCa-04 and miRCa-08 candidates, respectively. Relative expression was calculated normalizing the samples with the selected reference genes and the standard deviation resulted from three independent biological replicates.

The analysis of gene expression in reaction wood (RW_{2011}) model revealed different expression patterns for the three target genes. However, only punctual significant differences (P < 0.05) were observed between the tension and opposite wood samples in Eucgr.E01509 and Eucgr.C01382 target genes (Figure 19).

Expression of the Eucgr.E01509 target gene (miR171 candidate):

Considering the tension wood and opposite wood from one week of bending (P = 0.1961) and three/four weeks of bending (P = 0.1612), no significant decrease was revealed. However, by comparing the opposite and tension wood from one and three/four weeks of bending (P = 0.0317), we observed a slight increase of expression with statistically significance.

Expression of the Eucgr.C01382 target gene (miRCa-04 candidate):

This expression resulted in a statistically significant increase (P = 0.0385) between tension and opposite wood from one week of bending, followed by a no significant decrease (P = 0.2513) in the transition between tension and opposite wood from three/four weeks of bending.

Expression of the Eucgr.J02113 target gene (miRCa-08 candidate):

The expression of this target gene showed no significant differences in expression (P > 0.05) between the conditions under study.

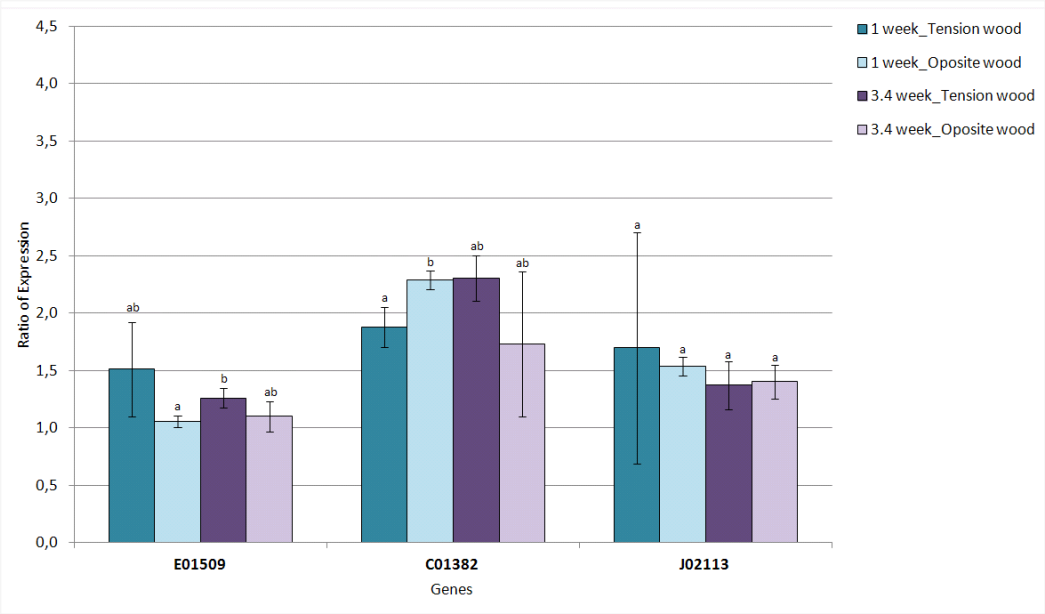


Figure 19: Quantitative real-time PCR (RT-qPCR) expression pattern of three *E. globulus* miRNAs target genes predicted and previously validated by RLM 5'-RACE methodology. Tension wood (dark blue) and opposite wood (light blue) providing from one week of bending, and tension wood (dark purple) and opposite wood (light purple) providing from junction of three/four week of bending are samples of developing xylem tissue, provided from reaction wood (RW_{2011}) model, respectively. Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113 are target genes of miR171, miRCa-04 and miRCa-08 candidates, respectively. Relative expression was calculated normalizing the samples with the selected reference genes and the standard deviation resulted from three independent biological replicates.

4

Discussion

Contents

4.1. Nucleic acids extraction in *Eucalyptus globulus*

Plants are diverse, and individual species organs or tissues of plants can behave differently during RNA extraction for use in molecular studies. The sample variations in this procedure include sampling techniques, the conservation, the presence of large quantities of polysaccharides, high levels of RNases, or different phenolic kinds (Chang *et al*., 1993). Therefore, it is essential to ensure a proper sample transport and storing coupled with basic cares to maintain RNA sample integrity during subsequent extraction.

The use of modified Provost *et al*. (2007) protocol enabled us to conduct successful total RNA extractions (see Figure 10, subsection §3.1.1.). Before the electrophoretic run in 2% (w/v) agarose gel, the application of a formamide denaturation of total RNA minimized the intramolecular interactions that could lead to bending of these molecules, changing the secondary structure of RNA and affecting the migration of the molecules in the gel (Kryndushkin *et al*., 2003). With this procedure, hydrogen bonding was broken and total RNA migrated as a single chain molecule (Ambion®, Applied Biosystems, 2012). The agarose gels [2% (w/v)] allowed a good separation of small fragments (Invitrogen®, Life Technologies, 2011). The addition of the SYBR Safe dye on gel allowed the visualization of bands on a transilluminator, which emits UV light. The inclusion of a convenient molecular marker was intended to establish a reference to assist us in a better estimate the size of RNA fragment. Total RNA is considered to be of poor quality and degraded when it reveals an indefinite smear of low molecular weight (Invitrogen®, Life Technologies, 2012). The integrity of the total RNA samples was exposed in the stained gels since our samples 28S and 18S bands appeared as sharp bands (Invitrogen®, Life Technologies, 2012).

The *NanoDrop ND-1000* spectrophotometer total RNA quantifications revealed considerable yield differences between samples. This is due to the fact that tissues under study came from different genotypes and environmental conditions. However, the purified RNA samples were considered good to proceed with the assay. Looking in detail to the $A_{260/280}$ and $A_{260/230}$ ratios (see Table 8, subsection §3.1.1.), we noticed different degrees of RNA contaminations (Gallagher and Desjardins, 2006). The $A_{260/280}$ ratio values attested the good RNA quality of the samples in terms of absence of phenolic and other contaminants (Thermo Scientific™, Wilmington, Delaware, USA). The ovaries and roots samples showed values below 2.00 but those values (1.98 and 1.83, respectively) were not considered problematic. The $A_{260/230}$ ratio values attested the polysaccharides presence in ovaries, pistils, roots and phloem samples. Polysaccharides tend to absorb at 230 nm (Becker *et al*., 2010). One the drawbacks associated to polysaccharides presence is the RNA mobility altered likely to cause distortion of the electrophoretic profile (see Figure 11, subsection §3.1.1.). These

results demonstrated the extremely sensitive nature of this biological material, emphasizing the difficulty in harvesting because of the high oxidation of plant material during sampling, the nature of the sample (richness in polysaccharides, e.g. stamens, ovaries and pistils), or because they are tissues/organs in contact with other components, such as soil or substrate (e.g. roots).

4.2. Discovery of miRNAs and their target genes in *Eucalyptus globulus*

The high level of sequence conservation of many mature miRNA sequences between distantly related species within the plant kingdom (Zhang *et al*., 2006) provides some homology search routes to identify conserved miRNAs from various plant species. In this study, computational prediction of miRNAs was conducted through bioinformatics tools. Computational prediction methods for miRNAs have obvious advantages, including the quick prediction of a large number of miRNAs, low costs, and the prediction of novel and non-abundant miRNAs that have been used in the identification of miRNAs in various plants, which are well dependent on the availability of ESTs and genomic sequences. The computational prediction of eucalypt miRNAs in this study is the first on record (Oliveira *et al*., 2013). Since precise sequences of miRNAs are essential for some important studies, such as miRNA target prediction, miRNA evolution, miRNA regulatory roles, and biogenesis mechanisms, the determination of the precise sequence of the potential miRNA is required.

The *CRAVELA* framework is a public pipeline developed for the analysis and presentation of data regarding the identification and evaluation of miRNA precursor candidates in metazoan genomes (Mendes *et al*., 2009; Mendes *et al*., 2010). Despite its original development context, the *CRAVELA* framework was successfully adapted to identify conserved miRNAs candidates in *Eucalyptus*. Among these candidates, we selected four eucalypt miRNAs candidates, which have been found to be homologous to conserved miRNA integrating four conserved MIR families: miRCa-02 candidate (MIR167), miRCa-04 candidate (MIR396), miRCa-08 candidate (MIR172) and miRCa-09 candidate (MIR477). Members of these families were previously identified in other species, such as *Arabidopsis* (Park *et al*., 2005; Rhoades *et al*., 2002; Palatnik *et al*., 2003; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004) or *Populus* (Lu *et al*., 2005; Zhang *et al*., 2006; Li *et al*., 2009).

4.2.1. Northern blot validation of selected miRNAs candidates

In this study, eight miRNAs candidates (guide and star* forms) were analyzed by Northern blot. The predicted size of miRNA was confirmed (20-24 bp) for each candidate, demonstrating that the bioinformatics analysis rendered the accurate predictions. MiRNAs

play a critical role in xylogenesis (Lu *et al*., 2005, Ong and Wickneswari, 2012), and also in response to biotic and abiotic stresses (Sunkar *et al*., 2012) and development (de Alba *et al*., 2012). The formation of tension wood (TW) and opposite wood (OW), collectively termed reaction wood (RW) (Pilate *et al*., 2004a; Pallardy, 2008) is a developmental feature characteristic of angiosperm trees and has been thought to be part of a stress-sensing mechanism leading to increased mechanical support (Paux *et al*., 2005; Qiu *et al*., 2008). On the other hand, the seasonal wood (SW) is one of the most environmental sources of variation affecting cambial activity and the formation of new cells, therefore influencing wood quality (Bedon *et al*., 2011). We examined miRNA transcript levels in two panels of tissues from eucalypt: the first, "developing xylem panel" includes exclusively eight different types of xylem tissues from seasonal wood (SW) and reaction wood (RW_{2009}) models, and mature wood; and the second, "other tissues/organs panel" includes total RNA from nine different tissues/organs. The quantification of guide and star* forms exhibited distinct tissue-specific expression patterns. The miRCa-02, miRCa-04 and miRCa-08 candidates almost always presented a higher expression level in seasonal wood model comparatively with reaction wood model (Figure 12). For example, in the miRCa-02 form, miRCa-04 (both forms) and miRCa-08* form, December (winter) is the month where we observe a decrease of expression, for the remaining three months of study. On the other hand, the miRCa-02 form and miRCa-04 form were found to be highly expressed in flowers (e.g. pollen), and the miRCa-08 form was more expressed in developing xylem (Figure 13). These observations suggest their functional association with cambium differentiation activities, but also with tissues and organs differentiation.

Both miRCa-02 (MIR167 family) forms presented a lowest expression level in developing xylem when compared with other tissues and organs (Figure 13). A similar result was observed in *Malus* where MIR167 was barely detectable in xylem (Varkonyi-Gasic *et al*., 2010). On the other hand, the miRCa-04 (MIR396 family) form showed a high expression level in pollen but also in developing xylem, leaves, roots and phloem tissues (Figure 13). Jones-Rhoades *et al*. (2004) reported in *Arabidopsis* that MIR396 presented the highest intensity of expression in leaves and roots. These results indicate that its activity in *Eucalyptus* seems to be similar to miRCa-02 (MIR167 family) and miRCa-04 (MIR396 family) in *Malus* and *Arabidopsis*, respectively. For the reaction wood model, miRCa-08 forms (MIR172 family), although not extensively, 09OW showed a higher expression level than 09TW (Figure 13). This result is interesting because Lu *et al*. (2005) observed in *Populus* that the expression of miR172 is suppressed in OW, reinforcing the regulatory role of this miRNA in different trees. On the other hand, the expression of MIR172 in *Malus* was barely detected in leaves, xylem and phloem (Varkonyi *et al*., 2010), but in *Eucalyptus* the miRCa-

08 (MIR172 family) showed the highest expression in developing xylem among the tissues and organs tested (Figure 13). All of these findings suggest that miRNA sequence conservation between plant species may not indicate conserved miRNA functions.

4.2.2. Experimental validation of miRNAs candidates and their target genes

The miRNAs generally function as negative regulators of gene expression by mediating the cleavage of target mRNAs (Llave *et al*., 2002b; Tang *et al*., 2003; Yekta *et al*., 2004; Song *et al*., 2010) or by repressing their translation (Chen, 2004; Gandikota *et al*., 2007; Brodersen *et al*., 2008; Lanet *et al*., 2009). Target prediction is an important step for assessment and assignment of putative functions to miRNAs in plants. Currently, the most efficient tool available for this is the bioinformatics approach since it is facilitated by the high degree of homology between miRNAs and their targets sequences in plants (Rhoades *et al*., 2002).

Target genes predictions were performed using the *psRNATarget* server for five selected miRNAs candidates (miR171, miRCa-02, miRCa-04, miRCa-08 and miRCa-09). A total of 68 predicted miRNAs-cleaved transcripts encoded by 42 genes were revealed. Target prediction was found to be the prevalent mechanism with 97% of the events being cases of miRNA-mediated target cleavage and only 3% by translation repression. These results were quite consistent with the literature, because according to some authors in plants the cleavage of target mRNAs appears to be predominant mode in gene regulation by miRNAs (Hutvager and Zamore, 2002; Kasschau *et al*., 2003; Carrington and Ambros, 2003; Ambros, 2003; Allison and Vaucheret, 2004; Bartel, 2004; Dugas and Bartel, 2004).

Finding the cleavage site supposedly located in the sequence complementary to the miRNA in the target gene is essential to validate the cleavage of target genes. One of the efficient methods to validate miRNAs by exposing miRNA dependent cleavage of targets is RNA ligase-mediated 5’ rapid amplification of cDNA ends, RLM 5’-RACE (Lu *et al*., 2005; Zhang *et al*., 2006; Song *et al*., 2012; Trindade, 2012). In total, we attempted RLM 5’-RACE miRNA validation on fourteen selected target genes, but only four produced positive results (Eucgr.E01509, Eucgr.E01875, Eucgr.C01382 and Eucgr.J02113). Nine target genes rendered negative results since the observed size of the amplification products did not corresponded to the expected products. Several reasons may account for this observation, such as the genomic differences between the *E. grandis* (640 Mbp, used for target prediction) and the *E. globulus* genomes (530 Mbp, used for target validation), or the possibility of low specificity of the PCR primers used. Moreover, it is known that computational tools for target prediction retrieve a high percentage of false positives for

miRNAs (Moxon *et al*., 2008; Thomson *et al*., 2011), which makes essential to conduct the experimental confirmation of these predictions (Mallory and Bouche, 2008). The results generated for Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113 genes confirmed the predicted cleavage site, attesting these genes as a miR171, miRCa-04 and miRCa-08 targets, respectively.

The *Arabidopsis* homologue of miR171, targets a member of the GRAS domain gene family, which was reported in to be involved in a variety of plant growth and developmental processes (Di Laurenzio *et al*., 1996; Reinhart *et al*., 2002; Mallory *et al*., 2005; Wang *et al*., 2005; Schommer *et al*., 2008). In *Populus*, the miR171 homologue has been described to present substantial expression in xylem formation (Lu *et al*., 2005). In our study, we confirmed the expression of this miRNA during developing xylem formation and validated Eucgr.E01509 gene as one of its targets (Figure 14). The RT-qPCR analysis showed that the expression of Eucgr.E01509 was significant different ($P < 0.05$) along the year from seasonal wood model (Figure 18), and also a decrease of its expression between opposite (OW) and tension (TW) tissues from reaction wood model (Figure 19). These observations suggest the involvement of this miRNA in wood formation under these environmental conditions. Reduced miR171 expression in mechanically stressed condition may have similar effects because its target genes are known to regulate cell division and elongation to produce cell lineages in response to gravitropism, a major mechanical stress for reaction wood induction (Di Laurenzio *et al*., 1996; Tasaka *et al*., 1999; Nakajima *et al*., 2001). For example, in poplar it was reported that miR171 expression in TW and OW are both similarly or preferentially altered between these tissues, contributing to control a set of coordinated pathways leading to the formation of reaction wood (Lu *et al*., 2004; Lu *et al*., 2005).

In *Arabidopsis* miRNAs from family MIR396 have been reported to target a member of the growth regulation (GRF) gene family that is involved in cell proliferation control during leaf and root development (Jones-Rhoades and Bartel 2004; Lu *et al*., 2007; Zhou *et al*., 2007). A homologue of this miRNA has been reported in *Medicago truncatula* to be cold stress-responsive (Devers *et al*., 2011; Trindade, 2012). Our study also confirmed that miRCa-04, the *E. globulus* miR396 homologue, was expressed during developing xylem differentiation, and targets the Eucgr.C01382 gene (Figure 16). Since, this gene has no known function (based on the bioinformatics prediction which relies on sequence homology against annotated genes from other species), it is possible that function of this miRNA it be specific to *Eucalyptus*. Despite this, Eucgr.C01382 target is a very interesting gene given its significant expression decrease in September, even if found not statistically significant. The expression of this gene also showed an inverse tendency in the reaction wood model (Figure 19), when we compared one week of bending to three/four weeks of bending, respectively.

Finally, miR172 was experimentally confirmed to target an APETALA2 (AP2) transcription factor coding gene. Homologues of this gene have been found to control many aspects of floral development and responses to the environment in species, such as *Arabidopsis* (Llave *et al*., 2002a; Park *et al*., 2005; Fahlgren *et al*., 2007), *Malus* (Xia *et al*., 2012) or poplar (Lu *et al*., 2005). In our work, we demonstrated that miRCa-08 (MIR172 family) targets Eucgr.J02113, a member of the *Eucalyptus* APETALA2 (AP2) transcription factor gene family that was expressed during xylem development (Figure 17). Contrasting results have been reported by Ong and Wickneswari (2012) in *Acacia mangium* since significant expression differences have been exposed between OW and TW tissues for miR172, whereas in our study no significant differences were revealed. The expression of Eucgr.J02113 was found to present a similar pattern to the ones presented by Eucgr.E01509 and Eucgr.C01382, corroborating its likely active on environmental responses during xylogenesis, as previously observed in other species (Laufs *et al*., 2004; Mallory *et al*., 2004; Sibout *et al*., 2008; Ong and Wickneswari, 2012). These results suggest that not only the sequences, but also the functions of miR171 and miRCa-08 miRNAs are conserved between *E. globulus* and other plant species.

Other interesting output of this RLM 5'-RACE validation work was the identification of new putative miRNA cleavage sites. In fact, we identified a non-predicted cleavage site for two target genes Eucgr.E01875 and Eucgr.C01382 (targeted by miRNAs miRCa-02 and miRCa-04, respectively). However, despite our extensive list of 3,300 miRNA candidates, it was not possible to predict regulator miRNAs, and new research is needed to identify these miRNAs. So, these results demonstrate that the experimental validation of predicted miRNA target genes provide us the ability to confirm their existence in eucalypt, but can also provide significant information on the roles of miRNAs in tree growth and development, including wood formation. Moreover, our experimental evidence suggests that expression of Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113 miRNAs target genes is noticeable responsive to environmental or mechanical stress.

5

Conclusion and Future Perspectives

Contents

5.1. Conclusion and future perspectives

This thesis contributes to the achievement of an important set of resources and associated knowledge on the post-transcriptional regulation occurring during the complex events underlying eucalypt xylogenesis. For the first time, four *E. globulus* miRNAs sequences and some of their target genes were disclosed and validated in parallel through bioinformatic and molecular tools.

The analysis of both guide and star* forms of four *E. globulus* miRNAs candidates (miRCa-02, miRCa-04, miRCa-08 and miRCa09) in an tissue sampling which includes developing xylem models and other tissues and organs proved the existence of miRNA tissue-specific post-transcriptional regulation of several genes. Thus, Northern blot allowed studying gene expression and confirmed its identity and involvement in the regulation of eucalyptus wood formation. Three potential target genes for three putative eucalypt miRNAs (miR171, miRCa-04 and miRca-08) were validated experimentally by RLM 5'-RACE, highlighting the role of miRNAs and their targets in *E. globulus* growth and development, particularly in xylogenesis. RLM 5'-RACE revealed to be an efficient and powerful approach that can be successfully used to validate the cleavage site of eucalypt miRNA targets. The multiplicity of *Eucalyptus* miRNAs (association with several distinct target genes) further suggests different roles of miRNAs in different tree species. The expression of some target genes was also evaluated by RT-qPCR, revealing different expression levels between developing xylem models. The three studied miRNAs, representatives of three conserved MIR families target genes responsible for plant growth, development, and stress response, were shown to be differentially expressed in eucalypt tissues. Some of the predicted miRNA targets were found to be in low abundance in tissues where miRNAs targets were detected. Predicted targets of miRNAs detected in the reaction wood model usually showed only minor differences in expression levels between tissues.

In order to complement this work, an extensive analysis from the Degradome will be conducted. Through this analysis we will try to understand the degree of accuracy obtained with the bioinformatic predictions of target genes, and also try to increase the experimental evidence of the present work, which for the first time allowed the validation of three genes as targets of three selected miRNAs in *E. globulus*. Given the amount of information provided by Degradome it will also be possible to identify a set of other target genes that are directly related to the formation of eucalyptus wood and relate this set of genes with the metabolic pathways in order to better understand the complex interactions exerted at transcriptional and post-transcriptional regulation levels during xylogenesis. Furthermore, the functional validation of these candidate genes will also be performed with transgenic plants over-expressing these genes and evaluating the phenotypic impact of the transformation.

References

Contents

6.1. References

Achard, P., Herr, A., Baulcombe, D., Harberd, N. (2004). Modulation of floral development by giberellin-regulated microRNAs. Dev. **131**, 3357-3365.

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., Sundaresan, V. (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res. **15**, 78–91.

Addo-Quaye, C., Eshoo, T., Bartel, D., Axtell, M. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. Curr. Biol. **18**, 758-762.

Alberts, B., Johnson, A., Lewis, J. Raff, M., Roberts, K., Walter, P. (2008). Molecular Biology of the Cell. New York: Garland Science. 5th edition, 538-539.

Alves-Junior, L., Niemeier, S., Hauenschild, A., Rehmsmeier, M., Merkle. T. (2009). Comprehensive prediction of novel microRNAs targets in *Arabidopsis thaliana*. Nucl. Acids Res. **37** (12), 4010.

Ambros, V., Lee, R., Lavanway, A., Williams, P., Jewell, D. (2003). MicroRNAs and other Tiny endogenous RNAs in *C. elegans*. Curr. Opin. Plant Biol. **13**, 807-818.

Anderson, C., Jensen, J., Orntoff, T. (2004). Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research. **64**, 5245–5250.

Arenas-Huertero, C., Perez, B., Rabanal, F., Blanco-Melo, D., Rosa, C., Estrada-Navarrete, G., Sanchez, F., Covarrubias, A., Reyes, J. (2009). Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. Plant Mol. Biol. **70**, 385-401.

Arend, M., Fromm, J. (2007). Seasonal change in the drought response of wood cell development in poplar. Tree Physiol. **27**, 985-992.

Aukerman, M., Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell. **15** (11), 2730-2741.

Baker, C., Sieber, P., Wellmer, F., Meyerowitz, E. (2005). The early extra petals 1 mutant uncovers a role for microRNAs miR164c in regulating petal number in *Arabidopsis*. Curr. Biol. **15**, 303-315.

Ballarin, A., Palma, H. (2003). Propriedades de resistência e rigidez da madeira juvenil e adulta de *Pinus taeda* L. Soc. Inv. Flores. **27** (3), 371-380.

Barnett, J., Bonham, V. (2004). Cellulose microfibril angle in the cell wall of wood fibres. Biol. Rev. Camb. Philos. Sci. **79**, 461-472.

Barrela, R. (2011). Identificação e sequenciação de genes da via da lenhificação em bibliotecas BAC de *Eucalyptus globulus*. Dissertação de Mestrado. Lisboa.

Bartel, B., Bartel, D. (2003). MicroRNAs: At the root of plant development. Plant Physiol. **132**, 709-717.

Bartel, D. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. **116**, 281-297.

Baumberger, N., Baulcombe, D. (2005). *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Nat. Acad. Sci. **102**(33), 11928-11933.

Becker, C., Fickinger-H, A., Riedmaier, I., Pfaffl, M. (2010). mRNA and microRNAs quality control for RT-qPCR analysis. Methods. **50**, 237-243.

Bedon, F., Grima-Pettenati, J., Mackay, J. (2007). Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). BMC Plant Biol. **7**, 17.

Bedon, F., Majada, J., Feito, I., Chaumeil, P., Dupuy, J., Lomenech, A., Barré, A., Gion, J., Plomion, C. (2011). Interaction between environmental factors affects the accumulation of root proteins in hydroponically grown *Eucalyptus globulus* (Labill.). Plant Physiol. **49**, 69-76.

Berezikov, E., Cuppen, E., Plasterk, R. (2006). Approaches to microRNA discovery. Nat. Genet. **38**, S2–S7.

Bernstein, E., Caudy, A., Hammond, S., Hannon, G. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. **409**, 363-366.

Boerjan, W., Ralph, J., Baucher, M. (2003). Lignin biosynthesis. Annu. Ver. Plant Biol. **54**, 519-546.

Bonnet, E., Wuyts, J., Rouze, P., Van de Peer, Y. (2004). Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc. Nat. Acad. Sci. **101**, 11511–11516.

Boudet, A., Kajita, S., Grima-Pettenati, J., Goffner, D. (2003). Lignins and lignocellulosics: a better control of synthesis for new and improved uses. Trends in Plant Sci. **8**, 576-581.

Bowman, J., Eshed, Y., Baum, S. (2002). Establishment of polarity in angiosperm lateral organs. Trends Genet. **18**(3), 134-141.

Brach, J., Frigerio, J-M., Claverol, S., Barré, A., Fevereiro, P., Plomion, C. (2008). Molecular and phenotypic profiling from the base to the crown in *maritime pine* wood-forming tissue. New Phytol. **178**, 283–301.

Brodersen, P., Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. Trends Genet. **22** (5), 268-280.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y., Sieburth, L., Voinnet, O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science. **320**, 1185–1190.

Brooker, M. (2000). A new classification of the genus *Eucalyptus* L'Hér. (*Myrtaceae*). Aust. Syst. Bot.**13**, 79–148.

Burton, R., Gibeaut, D., Bacic, A., Findlay, K., Roberts, K., Hamilton, A., Baulcombe, D., Fincher, G. (2000). Virus-induced silencing of a plant cellulose synthase gene. Plant Cell. **12**, 691-706.

Bustin, S. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endoc. **25**, 169-193.

Bustin, S. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol. Endocrinol. **29**, 23-39.

Carmell, M., Xuan, Z., Zhang, M., Hannon, G. (2002). The *Argonaute* family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. **16**, 2733-2733.

Carrington, J., Ambros, V. (2003) Role of microRNAs in plant and animal development. Science **301**, 336-338.

Carthew, R., Sontheimer, E. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell. **136**, 642-655.

Cassan-Wang, H., Soler, M., Yu, H., Camargo, E., Carocha, V., Ladouce, N., Savelli, B., Paiva, J., Grima-Pettenati, J. (2012). Reference genes for high throughput quantitative reverse transcription-PCR analysis of gene expression in organs and tissues of *Eucalyptus* grown in various environmental conditions. Plant Cell Physiol. **53** (12), 2101–16, 2012.

CELPA. (2011). Boletim Estatístico – Indústria Papeleira Portuguesa. Lisboa, pp. 19-30.

Chaffey, N., Cholewa, E., Regan, S., Sundberg, B. (2001). Secondary xylem development in *Arabidopsis*: a model for wood formation. Physiol Plant. 114, 594-600.

Chaffey, N. (2002). Why is there so little research into the cell biology of the secondary vascular system of trees? New Phytol. **153**, 213-223.

Chang, S., Puryear, J., Cairney, J. (1993). A simple and efficient method for isolation RNA from pine trees. Plant Mol. Biol. Rep. **11**, 113-116.

Chen, X., Liu, J., Cheng, Y., Jia, D. (2002). HEN1 functions pleiotropically in *Arabidopsis* development and acts in C function in the flower. Dev. **129**, 1085–1094.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. Science. **303**, 2022-2025.

Chen, X (2008). MicroRNA metabolism in plants. Curr.Top. Microbiol. Immunol. **320**, 117-136.

Chen, H., Chen, L., Patel, K., Li, Y., Baulcombe, D., Wu, S. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Nat. Acad. Sci. **107**, 15269-15274.

Christensen, J., Baucher, M., O'Connell, A., Van Montagu, M., Boerjan, W. (2000). Control of lignin biosynthesis. In S.M Jain and S.C Minocha (eds). Mol. Biol. of Woody Plants, Vol. **1**, 227-267.

Costa e Silva, F., Shvaleva, A., Maroco, J., Almeida, M., Chaves, M., Pereira, J. (2004). Responses to water stress in two *Eucalyptus globulus* clones differing in drought tolerance. Tree Physiology. **24**, 1165–1172.

Cuperus, J., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R., Takeda, A., Sullivan, C., Gilbert, S., Montgomery, T., Carrington, J. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. Nat. Mol. Biol. **17**, 997-1003.

Dai, X., Zhao, P. (2011). *psRNATarget*: a plant small RNA target analysis server. Nucl. Acids Res. **39**, W155-159.

Darley, C., Forrester, A., McQueen-Mason, S. (2001). The molecular basis of plant cell wall extension. Plant Mol. Biol. **47**, 179-195.

Debernardi, J., Rodriguez, R., Mecchia, M., Palatnik, J. (2012). Functional specialization of the plant miR396 regulatory network through distinct microRNA-target interactions. PLoS Genet. **8** (1), e1002419.

Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G., Leplé, J. (2010). Wood formation in Angiosperms. CR Biol. Apr. **333** (4), 325 – 334.

Demidenko, N., Logacheva, M., Penin, A. (2011). Selection and validation of reference genes for Quantitative Real-Time PCR in buckwheat *Fagopyrum esculentum* based on transcriptome sequence data. PLoS One, **6** (5):e19434.

Devers, E., Branscheid, A., May, P., Krajinski, F. (2011). Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. Plant Physiol. **156**, 1990-2010.

Dezulian, T., Palatnik, J., Huson, D., Weigel, D. (2006). Conservation and divergence of microRNA families in plants. Science. **313**, 1596-1604.

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M., Feldmann, K., Benfey,P. (1996). The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell **86**, 423-433.

Doench, J., Petersen, C., Sharp, P. (2003). siRNAs can function as miRNAs. Genes Dev. **17**, 438-442.

Druart, N., Johansson, A., Baba, K., Schrader, J., Sjödin, A., Bhalerao, T., Resman, L., Trygg, J., Moritz, T., Bhalerao, R. (2007). Environmental and hormonal regulation of the activity-dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. Plant Journal. **50**, 557-573.

Eamens, A., Wang, M., Smith, N., Waterhouse, P. (2008). NA silencing in plants: yesterday, today, and tomorrow. Plant Physiol. **147** (2), 456-468.

Eckardt, N. (2003). Cellulose synthesis takes the CesA train. Plant Cell **15**, 1685–1687.

Elbashir, S., Lendeckel, W., Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotides RNAs. Genes Dev. **15**, 188-200.

Emery, J., Floyd, S., Alvarez, J., Eshed, Y., Hawker, N., Izhaki, A., Baum, S., Bowman, J. (2003). Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. Curr. Biol. **13**, 1768-1774.

Farazi, T., Juranek, S., Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Dev. **135**, 1201-1214.

Foucart, C., Paux, E., Ladouce, N., San-Clemente, H., Grima-Pettenati, J., Sivadon, P. (2006). Transcript profiling of a xylem *vs* phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*. New Phytol. **170** (4),739–752.

Frazier, T., Sun, G., Burklew, C., Zhang, B. (2011). Salt and Drought Stresses Induce the Aberrant Expression of microRNA Genes in Tobacco. Mol Biotech. **49**, 159-165.

Gallagher, S., Desjardins, P. (2006). Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. Curr. Prot. Mol. Biol, Appendix 3: Appendix 3D.

Ghildiyal, M., Zamore, P. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. **10** (2), 94-108.

Gion, J., Carouché, A., Deweer, S., Bedon, F., Pichavant, F., Charpentier, J., Baillères, H., Rozenberg, P., Carocha, V., Ognouabi, N., Verhaegen, D., Grima-Pettenati, J., Vigneron, P., Plomion, C. (2011). Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. BMC Genomics. **12**, 301.

Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., Lapierre, C., Pollet, B., Verhaegen, D., Chaubet-Gigot, N., Grima-Pettenati, J. (2005). EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. Plant J. **43**(4), 553-567.

Goodstein, D., Shu, S., Howson, R., Neupane, R., Hayes, R. (2012). Phytozome: a comparative platform for green plant genomics. Nucl. Acids Res. **40** (D1), D1178–D1186.

Grattapaglia, D. (2008). *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytol. **179**, 911-929.

Grattapaglia, D., Kirst, M. (2008). *Eucalyptus* applied genomics: from gene sequences to breeding tools. New Phytol. 179 (4), 911–29.

Grattapaglia, D., Plomion, C., Kirst, M., Sederoff, R. (2009). Genomics of growth traits on forest trees. Curr. Opin. Gen. Dev. **12** (2), 148-156.

Grennan, A. (2008). *Arabidopsis* microRNAs. Plant Physiol. **146**, 3-4.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucl. Acids Res. **32**, 109-111.

Griffiths-Jones, S., Grocock, R., Van Dongen, S., Bateman, A., Enright, A. (2006). MiRBase: microRNA sequences, targets and gene nomenclature. Nucl. Acids Res. **34**, 140-144.

Griffiths-Jones, S., Saini, H., Van Dongen, S., Enright, A. (2008). miRBase: tools for microRNA genomics. Nucl. Acids Res. **36**, 154–8.

Grima-Pettenati, J., Goffner, D. (1999). Lignin genetic engineering revisited. Plant Sci. **145**, 51-65.

Guo, Y., Jangi, S., Welte, M. (2005). Organelle-specific control of intracellular transport: distinctly targeted isoforms of the regulator Klar. Mol. Biol. Cell **16** (3), 1406-1416.

Hagen, G., Guilfoyle, T. (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol. Biol. **49**, 373-385.

Hall, T. (1999). *BioEdit*: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Res. **41**, 95-98.

Hamilton, A., Baulcombe, D. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286**, 950-952.

Hamilton, M., Raymond, C., Harwood, C., Potts, B. (2009). Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. Genet. Genom. **5** (2), 307-316.

Han, K. (2001). Molecular biology of secondary growth, Journal of Plant Biotechnology **3**, 45-57.

Han, M., Goud, S., Song, L., Fedoroff, N. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc. Nat. Acad. Sci. **101**, 1093-1098.

Hertzberg, M., Aspeborg, H., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlen, M., Teeri, T., Lundeberg, J., Sundberg, B., Nilsson, P., Sandberg, G. (2001). A transcriptional roadmap to wood formation. Proc. Nat. Acad. Sci. **98**, 14732-14737.

Higuchi, R., Dollinger, G., Walsh, P., Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. Biotechnology, **10**, 413–417.

Hirayama, T., Shinozaki, K. (2010). Research on plant abiotic stress responses in the postgenome era: past, present and future. Plant Journal. **61**, 1041-1052.

Jack, T. (2004). Molecular and genetic mechanisms of floral control. Plant Cell **16**, S1-S17.

Jaillon, O., Aury, J., Noel. B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. **449**, 463-467.

Jones, A. (2001). Programmed cell death in development and defense. Plant Physiol. **125**, 94-97.

Jones-Rhoades, M., Bartel, D. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell. **14**(6), 787-799.

Jones-Rhoades, M., Bartel, D., Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. **57**, 19-53.

Jourez, B., Riboux, A., Leclercq, A. (2001) Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* CV ''Ghoy''). IAWA J **22**, 133–157.

Juarez, M., Kui, J., Thomas, J., Heller, B., Timmermans, M. (2004). microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. Nature **428**, 84-88.

Kasschau, K., Xie, Z., Allen, E., Llave, C., Chapman, E., Krizan, K., Carrington, J. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev. Cell. **4**, 205-217.

Kidner, C., Martienssen, R. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature **428**, 81-84.

Kim, J., Jung, J., Reyes, J., Kim, Y., Kim, S., Chung, K., Kim, J., Lee, M., Lee, Y., Narry, K., Chua, N., Park, C. (2005). MicroRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems. Plant J **42**, 84-94.

Kozomara, A., Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucl. Acids Res. **39**, 152-157.

Kurihara, Y., Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. **101**, 12753–12758.

Kurihara, Y., Takashi, Y., Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. RNA. **12**, 206–212.

Ladiges, P., Udovicic, F., Nelson, G. (2003). Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. J. Biogeog. **30**, 989-998.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science. **294**, 853-858.

Laufs, P., Peaucelle, A., Morin, H., Traas, J. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. Development **131**, 4311-4322.

Lee, R., Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science. **294**, 862-864.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature **425**, 415-419.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO Journal. **23**(20), 4051-4060.

Legay, S., Lacombe, E., Goicoechea, M., Briere, C., Seguin, A., Mackay, J., Grima-Pettenatti, J. (2007). Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. Plant Sci. **173**, 542-549.

Leonello, E., Palma, H., Ballarin, A. (2008). Delimitação da madeira juvenil e adulta de *Eucalyptus grandis* em São Paulo, Brasil. Rev. Flor. Venez. **52** (1) : 93-98.

Li, J., Yang, Z., Yu, B., Liu, J., Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3' end uridylation activity in *Arabidopsis*. Curr. Opin. Plant Biol. **15**, 1501–1507.

Li, B., Yin, W., Xia, X. (2009). Identification of microRNAs and their targets from *Populus euphratica*. Biochem. Biophys. Res. Commin. **388**, 272-277.

Llave, C., Xie, Z., Kasschau, K., Carrington, J. (2002a). Cleavage of scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science. **297**, 2053–2056.

Llave, C., Xie, Z., Kasschau, K., Carrington, J. (2002b).Endogenous and silencing-associated small RNAs in plants. Plant Cell. **14**, 1605-1619.

Lobbes, D., Rallapalli, G., Schmidt, D., Martin, C., Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. EMBO Rep **7**, 1052-1058.

Lohmann, J., Weigel, D. (2002). Building beauty : the genetic control of floral patterning. Dev. Cell **2**, 135-142.

Lu, S., Sun, Y-H., Shi, R., Clark, C., Li, L., Chiang, V. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. Plant Cell. **17**, 2186-2203.

Lu, C., Kulkarni, K., Souret, F., MuthuValliappan, R., Tej, S., Poethig, R., Henderson, I., Jacobsen, S., Wang, W., Green, P., Meyers, B. (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. Genome Res. **16**, 1276-1288.

Lu, S., Sun,Y-H., Amerson, H., Chiang, V. (2007). MicroRNAs in *loblolly pine* (*Pinus taeda* L.) and their association with fusiform rust gall development. Plant J. **51**, 1077-1098.

Lv, D., Bai, X., Li, Y., Ding, X., Ge, Y., Cai, H., Ji, W., Wu, N., Zhu, Y. (2010). Profiling of cold-stress-responsive miRNAs in rice by microarrays. Gene **459**, 39-47.

Magaton, A., Oliveira, R., Lopes, O., Lilagres, F., Piló-Velose, D., Colodette, J. (2000). Composição Quimica da Madeira da espécie de Eucalipto. 29ª Reunião Anual da Sociedade de Química Brasileira. Instituto de Química da UPS. São Paulo-SP-Brasil.

Mallory, A., Dugas, D., Bartel, D., Bartel, B. (2004). MicroRNA regulation of NAC domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr. Biol. **14**, 1035-1046.

Mallory, A., Bartel, D., Bartel, B. (2005). MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell **17**, 1360-1375.

Mallory, A., Elmayan, T., Vaucheret, H. (2008) MicroRNA maturation and action - the expanding roles of ARGONAUTEs. Curr. Opin. Plant Biol, **11** (5), 560–6.

Mallory, A., Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. Plant Cell **22**, 3879-3889.

Mandrou, E., Mortier, F., Denis, M., Chaix, G., Villar, E., Plomion, C., Gion, J-M. (2011). Functional variability of two lignification genes in *Eucalyptus urophylla*. BMC Proc. **5** (7), 12.

Marques, C., Brondani, R., Grattapaglia, D., Sederoff, R. (2002). Conservation and synteny of SSR loci and QTLs for vegetative propagation in four Eucalyptus species. Theor. App. Genet., **105** (2-3), 474–478.

Mellerowicz, E., Baucher, M., Sundberg, B., Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. Plant Mol. Biol. **47**, 239-274.

Mendes, N., Freitas, A., Sagot, M. (2009). Current tools for the identification of miRNA genes and their targets. Nucl. Aci. Res., **37** (8), 2419–33.

Mendes, N., Freitas, A., Vasconcelos, A., Sagot, M. (2010). Combination of measures distinguishes pre-miRNAs from other stem-loops in the genome of the newly sequenced Anopheles darling. BMC Genomics. **11**, 529**.**

Merchan, F., Boualem, A., Crespi, M., Frugier, F. (2009). Plant polycistronic precursors containing non-homologous microRNAs target transcripts enconding functionally related proteins. Genome Biol. **10**, 136.

Mette, M., Aufsatz, W., Van der Winder, J., Matzke, M., Matzke, A. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBRO J. **19**, 5194-5201.

Meyers, B., Axtell, M., Bartel, B., Bartel, D., Baulcombe, D., Bowman, J., Cao, X., Carrington, J., Chen, X., Green, P., Griffiths-Jones, S., Jacobsen, S., Mallory, A., Martienssen, R., Poethig, R., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D., Zhu, J. (2008). Criteria for annotation of plant MicroRNAs. Plant Cell **20**, 3186-3190.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G., Qi, Y. (2008). Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. Cell **133**, 116-127.

Mica, E., Piccolo, V., Delledonne, M., Ferrarini, A., Pezzotti, M., Casati, C., Del-Fabbro, C., Valle, G., Policriti, A., Morgante, M., Pesole, G., Pe, M., Horner, D. (2009). High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in *Vitis vinifera*. BMC Genom. **10**, 58.

Millar, A., Waterhouse, P. (2005). Plant and animal microRNAs: similarities and differences. Funct. Integ. Genom., **5** (3),129-135.

Myburg, A., Grattapaglia, D., Tuskan, G., Hellsten, U., Hayes, R., Grimwood, J., *et al*. (2013). Genome sequence of *Eucalyptus grandis*: a global tree crop for fiber and energy. Submitted.

Mizrachi, E., Hefer, C., Ranik, M., Joubert, F., Myburg, A. (2010). *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by *Illumina mRNA-Seq*. BMC Genom., **11** (1), 681.

Moldovan, D., Spriggs, A., Yang, J., Pogson, B., Dennis, E., Wilson, I. (2010). Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in *Arabidopsis*. J Exp. Bot. **61**,165-177.

Moxon, S., Schwach, F., Dalmay, T., MacLean, D., Studholme, D., Moulton, V. (2008). A toolkit for analyzing large-scale plant small RNA datasets. Bioinf. App. Note. **24**, 2252-2253.

Moyle, R., Schrader, J., Stenberg, A., Olsson, O., Saxena, S., Sandberg, G., Bhalerao, R. (2002). Environmental and auxin regulation of wood formation involves members of the Aux/IAA gene family in hybrid aspen. Plant Journal. **31**, 675-685.

Nascimento, L., Neto, J., Salaza, M., Camargo, E., Marques, W., Gonçalves, D., Vidal, R., Pereira, G., Carazzolle, M. (2011). Na integrated databese of *Eucalyptus spp*. genome project. BMC Proc. **5** (7), 170.

Neale, D. (2007). Genomics to tree breeding and forest health. Curr. Opin. Gen. Dev. **17** (6), 539-544.

Novaes, E., Drost, D., Farmerie, W., Pappas, G., Grattapaglia, D., Sederoff, R., Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. BMC Genomics. **9**, 312.

Oliveira, J., Mendes, N., Carocha, V., Graça, C., Paiva, J., Freitas, A. (2013). A computational approach for microRNA identification in plants: combining genome-based predictions with RNA-Seq data. Genom. Proteom. **4** (2), 1000130.

Ong, S., Wickneswari, R. (2012). Characterization of miRNAs expressed during secondary wall biosynthesis in *Acacia mangium*. PLoS one **7** (11), e49662.

Paiva, J., Garcés, M., Alves, A., Garnier-Géré, P., Rodrigues, J., Lalanne, C., Porcon, S., Provost, G., Perez, D., Brach, J., Frigerio, J-M., Claverol, S., Barrés, A., Fevereiro, P., Plomion, C. (2008). Molecular and phenotypic profiling from the base to the crown in *maritime pine* wood-forming tissue. New Phytol. **178**, 283-301.

Paiva, J., Prat, E., Vautrin, S., Santos, M., San-Clemente, H., Brommonschenkel, S., Fonseca, P., Grattapaglia, D., Song, X., Ammiraju, J., Kudrna, D., Wing, R., Freitas, A., Bergès, H., Grima-Pettenati, J. (2011). Advancing *Eucalyptus* genomics: identification and sequencing of lignin biosynthesis genes from deep-coverage BAC libraries. BMC Genom., **12** (1), 137.

Palatnik, J., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J., Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature **425**, 257-263.

Pall, G., Codony-Servat, C., Byrne, J., Ritchie, L., Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. Nucl. Acids Res **35**, e60.

Pallardy S. (2008). The woody plant body. In Physiology of Woody Plants (Chapter 2, Third Edition), pages 9 – 38. Academic Press, San Diego.

Pantaleo, V., Szittya, G., Moxon, S., Miozzi, L., Moulton, V., Dalmay, T., Burgyan, J. (2010). Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. Plant Journal. **62**, 960-976.

Papp, I., Mette, M., Aufsatz, W., Daxinger, L., Schauer, S., Ray, A., Winden, J., Matzke, M., Matzke, A. (2003). Evidence for nuclear processing of plant microRNA and short interfering RNA precursors. Plant Physiol. **132**, 1382-1390.

Parsons, M., Gavran, M., Davidson, J. (2006). Australia's Plantations. Department of Agriculture, Fisheries and Forestry, Canberra, 1-56.

Park, M., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R. (2005). Nuclear processing and export of microRNAs in *Arabidopsis*. Proc. Natl. Acad. Sci. **102**, 3691-3696.

Paux, E., Tamasloukht, M'B., Ladouce, N., Sivadon, P., Grima-Pettenati, J. (2004). Identification of genes preferentially expressed during wood formation in *Eucalyptus*. Plant Mol. Biol. **55**, 263-280.

Paux, E., Carocha, V., Marques, C., Sousa, A., Borralho, N., Sivadon, P., Grima-Pettenati, J. (2005). Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. New Phytol. **167**, 89-10.

Pereira, J. (2007). Uma espécie altamente produtiva. *In* Pinhais e Eucaliptais - A floresta cultivada. Lisboa: Público SA, Fundação Luso-Americana para o Desenvolvimento. 167-177.

Pereira, H., Miranda, I., Gominho, J., Tavares, F., Quilhó, T., Graça, J., Rodrigues, J., Shatalov, A., Knapic, S. (2010). Qualidade tecnológica do eucalipto *Eucalyptus globulus*. *In* Centro de Estudos Florestais (Ed.), ISA-UTL, Lisboa.

Pfaffl, M. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucl. Aci. Res. **29**, 9.

Pilate, G., Dejardin, A., Laurans, F., Leple, J-C. (2004a). Tension wood as a model for functional genomics of wood formation. New Phytol. **164**, 63-72.

Pilate, G., Chabbert, B., Cathala, B., Yoshinaga, A., Leplé, J-C., Laurans, F., Lapierre, C., Ruel, K. (2004b). Lignification and tension wood. CR. Biol. **327**, 889-901.

Pinho, J. (2010). 5.º Inventário Florestal Nacional - Apresentação do Relatório Final. SEFDR.

Pita, P., Pardos, J. (2001). Growth, leaf morphology, water use and tissue water relations of *Eucalyptus globulus* clones in response to water deficit. Tree Physiol. **21**, 599-607.

Plomion, C., Pionneau, C., Brach, J., Costa, P., Bailleres, H. (2000). Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol. **123**, 959-969.

Plomion, C., Provost, G., Stokes, A. (2001). Wood Formation in Trees. Plant Physiol. **127**, 1513-1523.

Poke, F., Vaillancourt, R., Potts, B., Reid, J. (2005). Genomic research in *Eucalyptus*. Genetica. **125**, 79-101.

Potts, B., Dungey, H. (2004). Interspecific hybrids of *Eucalyptus*: key issues for breeders and geneticists. New Forests. **27**(2), 115-138.

Provost, G., Herrera, R., Paiva, J., Chaumeil, P., Salin, F., Plomion, C. (2007). A micromethod for high throughput RNA extraction in forest trees. Biol. Res. **40**, 291-297.

Qiu, D., Wilson, I., Gan, S., Washusen, R., Moran, G., Southerton, S. (2008). Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. New Phytol. **179**, 94-103.

Rajagopalan, R., Vaucheret, H., Trejo, J., Bartel, D. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev **20**, 3407-3425.

Ranik, M., Creux, N., Myburg, A. (2006). Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. Tree Physiol. **26**, 365-375.

Raymond, C. (2002). Genetics of *Eucalyptus* wood properties. Ann. For. Sci. **59**, 525–531.

Reinhart, B., Weinstein, E., Rhoades, M., Bartel, B., Bartel, D. (2002). MicroRNAs in plants. Genes Dev. **16**, 1616-1626.

Rengel, D., San-Clemente, H., Servant, F., Ladouce, N., Paux, E., Wincker, P., Couloux, A., Sivadon, P., Grima-Pettenati, J. (2009). A new genomic resource dedicated to wood formation in *Eucalyptus*. BMC Plant Biol. **9**, 36.

Rhoades, M., Reinhart, B., Lim, L., Burge, C., Bartel, B., Bartel, D. (2002). Prediction of plant microRNA targets. Cell. **110**:513-520.

Roberts, K., McCann, M. (2000). Xylogenesis: the birth of a corpse. Curr. Opin. Plant Biol. **3**, 517-522.

Rodrigues, J., Graça, J., Pereira, H. (2000). Influence of tree eccentric growth on syringyl/guaiacyl ratio in *Eucalyptus globulus* wood lignin assessed by analytical pyrolysis. J. Anal. App. Pyrol. **58**, 481-489.

Rodriguez, R., Mecchia, M., Debernardi, J., Schommer, C., Weigel, D., Palatnik, J. (2010). Control of cell proliferation in *Arabidopsis thaliana* by microRNA miR396. Development **137**, 103-112.

Ruiz-Medrano, R., Xoconostle-Cázares, B., Lucas, W. (2001). The phloem as a conduit for inter-organ communication, Current Opinion in Plant Biology **4**, 202.

Salazar, M., Nascimento, L., Camargo, E., Goncalves, D., Neto, J., Marques, W., Teixeira, P., Mieczkowski, P., Mondego, J., Carazzolle, M., Deckmann, A., Pereira, G. (2013). Xylem transcription profiles indicate potential metabolic responses for economically relevant characteristics of *Eucalyptus* species. BMC Genom., **14** (1), 201.

Schlamp, K., Weinmann, A., Krupp, M., Maass, T., Galle, P., Teufel, A. (2008). BlotBase: A Northern blot database. Gene, **427**, 47–50.

Schneider, C., Rasband, W., Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Met **9**, 671-675.

Schommer, C., Palatnik, J., Aggarwal, P., Chetelat, A., Cubas, P., Farmer, E., Nath, U., Weigel, D. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. PLoS Biol **6**, e230.

Schwab, R., Palatnik, J., Riester, M., Schommer, C., Schmid, M., Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev. Cell **8**, 517 – 527.

Schwab, R., Maizel, A., Ruiz-Ferrer, V., Garcia, D., Bayer, M., Crespi, M., Voinnet, O., Martienssen, R. (2009). Endogenous TasiRNAs mediate non-cell autonomous effects on gene regulation in *Arabidopsis thaliana*. PLoS One **4**, e5980.

Schwarz, S., Grande, A., Bujdoso, N., Saedler, H., Huijser, P. (2008). The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol Biol **67:** 183-195.

Silva, J. (2007). Àrvores e Florestas de Portugal. In Pinhais e Eucaliptais - A floresta Cultivada. Lisboa: Público, SA, Fundação Luso-Americana para o Desenvolvimento. 149-150.

Shivaprasad, P., Chen, H., Patel, K., Bond, D., Santos, B., Baulcombe, D. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell **24**, 859-874.

Song, C., Fang, J., Li, X., Liu, H., Thomas, C. (2009). Identification and characterization of 27 conserved microRNAs in C*itrus*. Plant. **230**, 671-685.

Sunkar, R., Zhu, J. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell. **16**, 2001-2019.

Sunkar, R., Li, Y-F., Jagadeeswaran, G. (2012). Functions of miRNAs in plant stress responses. Trends Plant Sci. **17** (4),196-203.

Szittya, G., Moxon, S., Santos, D., Jing, R., Fevereiro, M., Moulton, V., Dalmay, T. (2008). High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNAs families. BMC Genomics. **9**, 593.

Tang, G., Reinhart, B., Bartel, D., Zamore, P. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. **17**, 49-63.

Thamarus, K., Groom, K., Bradley, A., Raymond, C., Schimleck, L., Williams, E., Moran, G. (2004). Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of Eucalyptus globulus. Theor. Appl. Genet. **109**, 856-864.

Thumma, B., Southerton, S., Bell, J., Owen, J. Henery, M., Moran, G. (2010). Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens.* Tree Genet Genomes. **6**, 305–317.

Trindade, I. (2012). MicroRNAs as modulators of water deficit responses in *Medicago truncatula*. Dissertation of Ph.D degree, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

Tuskan, G., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R., Bhalerao, R., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapm, C. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. **313**, 1596-1604.

Uggla, C., Magel, E., Moritz, T., Sundberg, B. (2001). Function and dynamics of auxin and carbohydrates during earlywood/latewood transition in Scots pine. Plant Physiol. **125**, 2029-2039.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., Rozen, S. (2012). Primer3-new capabilities and interfaces. Nucl. Acids Res. **40** (15), e115.

Unver, T., Namuth-Covert, D., Budak, H. (2009). Review of Current Methodological Approaches for Characterizing MicroRNAs in Plants. Int. J. Plant Genomics. **262463**, 11.

Vaucheret, H., Vazquez, F., Crete, P., Bartel, D. (2004). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev. **18**, 1187-1197.

Vazquez, F., Blevins, T., Ailhas, J., Boller, T., Meins, F. (2008). Evolution of *Arabidopsis* MIR genes generates novel microRNA classes. Nucl. Acids Res. **36**, 6429-6438.

Villar, E., Klopp, C., Noirot, C., Novaes, E., Kirst, M., Plomion, C., Gion, J. (2011). RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. BMC Genom. **12**, 538.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell. **136**, 669-687.

Wan, L., Wang, F., Guo, X., Lu, S., Qiu, Z., Zhao, Y., Zhang, H., Lin, J. (2012). Identification and characterization of small non-coding RNAs from Chinese fir by high throughput sequencing. BMC Plant Biol. **12**,146.

Wang, J., Wang, L., Mao, Y., Cai, W., Xue, H., Chen, X. (2005). Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. Plant Cell **17**, 2204-2216.

Weijers, D., Jurgens, G. (2005). Auxin and embryo axis formation: the ends in sight?. Curr. Opin. Plant Biol. **8**, 32-37.

Williams, L., Grigg, S., Xie, M., Christensen, S., Fletcher, C. (2005). Regulation of *Arabidopsis* shoot apical meristem and lateral organ formation by microRNA *miR166g* and its AtHD-ZIP target genes. Dev. **132**, 3657-3668.

Wu, G., Poethig, R. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. Development **133**, 3539-3547.

Xia, R., Zhu, H., An, Y., Beers, E., Liu, Z. (2012). Apple miRNAs and tasiRNAs with novel regulatory networks. Gen. Biol. **13**, R47.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S., Carrington, J. (2005). Expression of *Arabidopsis* MIRNA genes. Plant Physiol. **138**, 2145-2154.

Xu, Q., Liu, Y., Zhu, A., Wu, X., Yu, K., Guo, W., Deng, X. (2010). Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. BMC Genomics. **11**, 246

Yakovlev, I., Fossdal, C., Johnsen, O. (2010). MicroRNAs, the epigenetic memory and climatic adaptation in Norway spruce. New Phytol. **187**, 1154-1169.

Yamaguchi-Shinozaki, K., Shinozaki, K. (2005). Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci **10**, 88-94.

Yan, K., Yan, S., Farooq, A., Han, A., Zeng, L., Zhou, M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature. **426**, 468-474.

Yang, Z., Ebright, Y., Yu, B., Chen, X. (2006). HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 20 OH of the 30 terminal nucleotide. Nucl. Acids Res. **34**, 667–675.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R., Steward, R., Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science **307**, 932–935.

Zhang, B., Pan, X., Cobb, G., Anderson, T. (2006). Plant microRNA: A small regulatory molecule with big impact. Dev. Biol. **289**, 3-16.

Zeng, C., Wang, W., Zheng, Y., Chen, X., Bo, W., Song, S., Zhang, W., Peng, M. (2010). Conservation and divergence of microRNAs and their functions in *Euphorbiaceous* plants. Nucl. Acids Res. **38**:981-995.

Zhong, R., Ye, Z-H. (2007). Regulation of cell wall biosynthesis. Curr. Opin.Plant Biol. **10**, 564-572.

Zhou, X., Wang, G., Zhang, W. (2007). UV-B responsive microRNA genes in *Arabidopsis thaliana*. Mol. Syst. Biol., **3**, 103.

Zobel, B., Buijtenen, J. (1989). Wood variation: its causes and control. Springer-Verlag, Berlin, Germany.

http://www.fasteris.com

http://www.gene-quantification.de/download.html

http://www.phytozome.net/cgi-bin/gbrowse/eucalyptus/)

http://www.mirbase.org/cgi-bin/browse.pl

http://genomique.genotoul.fr/

http://acgt.cs.tau.ac.il/expander/

www.cravela.org

http://www.quantprime.de/

http://bioinformatics.cau.edu.cn/PMDR/

Appendix
Tables and Figures

Appendix I: Supplementary Tables

Suplementary Table S1a. List of *in silico* prediction of miRNAs target genes obtained using the *psRNATarget* software (Xinbin *et al*., 2011).

MiRNAs Candidates	Precursor Name	Target Gene	Target Transcript	Target Gene Annotation	Expectation	Inhibition	Multiplicity
miRCa-02	mirC-6538426	Eucgr.E01875	Eucgr.E01875.1	Unknown	0.5	Cleavage	1
		Eucgr.K00370	Eucgr.K00370.1	CCT motif family protein	2.5	Cleavage	1
		Eucgr.K00370	Eucgr.K00370.2	CCT motif family protein	2.5	Cleavage	1
		Eucgr.K00370	Eucgr.K00370.3	CCT motif family protein	2.5	Cleavage	1
		Eucgr.K00370	Eucgr.K00370.4	CCT motif family protein	2.5	Cleavage	1
		Eucgr.B01226	Eucgr.B01226.1	Basic helix-loop-helix (bhlh) DNA-binding superfamily protein	2.5	Cleavage	1
		Eucgr.D02645	Eucgr.D02645.1	Homeobox 7	3.0	Cleavage	1
miRCa-04	mirC-7274614	Eucgr.B02279	Eucgr.B02279.1	RNA helicase, ATP-dependent, SK12/DOB1 protein	2.0	Cleavage	1
		Eucgr.C01382	Eucgr.C01382.1	Unknown	2.0	Cleavage	1
		Eucgr.D01126	Eucgr.D01126.1	Seven transmembrane MLO family protein	2.5	Cleavage	1
		Eucgr.D01126	Eucgr.D01126.2	Seven transmembrane MLO family protein	2.5	Cleavage	1
		Eucgr.L01672	Eucgr.L01672.1	Seven transmembrane MLO family protein	3.0	Cleavage	1
		Eucgr.D01758	Eucgr.D01758.1	Ftsh extracellular protease family	3.0	Cleavage	1
		Eucgr.D01758	Eucgr.D01758.2	Ftsh extracellular protease family	3.0	Cleavage	1
		Eucgr.D01758	Eucgr.D01758.3	Ftsh extracellular protease family	3.0	Cleavage	1
		Eucgr.D01758	Eucgr.D01758.4	Ftsh extracellular protease family	3.0	Cleavage	1
		Eucgr.E03139	Eucgr.E03139.1	Unknown	3.0	Cleavage	1
		Eucgr.H02289	Eucgr.H02289.1	Protein of unknown function (DUF594)	3.0	Cleavage	1
miRCa-08	mirC-8742859	Eucgr.J02113	Eucgr.J02113.1	Related to AP2.7	0.5	Cleavage	1
		Eucgr.J02113	Eucgr.J02113.2	Related to AP2.7	0.5	Cleavage	1
		Eucgr.A01182	Eucgr.A01182.1	Related to AP2.7	1.5	Cleavage	1
		Eucgr.G02793	Eucgr.G02793.1	Related to AP2.7	1.5	Cleavage	1
		Eucgr.G02793	Eucgr.G02793.2	Related to AP2.7	1.5	Cleavage	1
		Eucgr.B02453	Eucgr.B02453.1	Related to AP2.7	2.0	Cleavage	1

Suplementary Table S1b. List of *in silico* prediction of miRNAs target genes obtained using the *psRNATarget* software (Xinbin *et al*., 2011; Continuation).

MiRNAs Candidates	Precursor Name	Target Gene	Target Transcript	Target Gene Annotation	Expectation	Inhibition	Multiplicity
miRCa-08	mirC-8742859	Eucgr.B02453	Eucgr.B02453.2	Related to AP2.7	2.0	Cleavage	1
(Continuation)		Eucgr.I00892	Eucgr.I00892.1	Integrase-type DNA-binding superfamily protein	2.0	Cleavage	1
		Eucgr.I00892	Eucgr.I00892.2	Integrase-type DNA-binding superfamily protein	2.0	Cleavage	1
		Eucgr.I00892	Eucgr.I00892.3	Integrase-type DNA-binding superfamily protein	2.0	Cleavage	1
		Eucgr.J00901	Eucgr.J00901.3	Unknown	2.0	Cleavage	1
		Eucgr.J00901	Eucgr.J00901.2	Unknown	2.0	Cleavage	1
		Eucgr.J00901	Eucgr.J00901.1	Unknown	2.0	Cleavage	1
		Eucgr.J00901	Eucgr.J00901.4	Unknown	2.0	Cleavage	1
		Eucgr.J00901	Eucgr.J00901.5	Unknown	2.0	Cleavage	1
		Eucgr.E03545	Eucgr.E03545.1	Unknown	2.0	Cleavage	1
		Eucgr.J01237	Eucgr.J01237.1	Unknown	2.5	Cleavage	1
		Eucgr.I00713	Eucgr.I00713.1	Protein kinase superfamily protein	2.5	Cleavage	1
		Eucgr.I00713	Eucgr.I00713.2	Protein kinase superfamily protein	2.5	Cleavage	1
		Eucgr.I00713	Eucgr.I00713.3	Protein kinase superfamily protein	2.5	Cleavage	1
		Eucgr.I00713	Eucgr.I00713.4	Protein kinase superfamily protein	2.5	Cleavage	1
		Eucgr.K01320	Eucgr.K01320.1	GRAS family transcription factor	2.5	Cleavage	1
		Eucgr.J03088	Eucgr.J03088.1	Sterile alpha motif (SAM) domain-containing protein	3.0	Cleavage	1
		Eucgr.I02510	Eucgr.I02510.2	D-aminoacid aminotransferase-like PLP-dependent enzymes	3.0	Cleavage	1
		Eucgr.I02510	Eucgr.I02510.3	D-aminoacid aminotransferase-like PLP-dependent enzymes	3.0	Cleavage	1
miRCa-09	mirC-8853075	Eucgr.H03865	Eucgr.H03865.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.H03863	Eucgr.H03863.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.L02566	Eucgr.L02566.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.H03864	Eucgr.H03864.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.L01962	Eucgr.L01962.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.D01892	Eucgr.D01892.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.D01900	Eucgr.D01900.1	Osmotin 34	2.5	Cleavage	1

Suplementary Table S1c. List of *in silico* prediction of miRNAs target genes obtained using the *psRNATarget* software (Xinbin *et al*., 2011; Continuation).

MiRNAs Candidates	Precursor Name	Target Gene	Target Transcript	Target Gene Annotation	Expectation	Inhibition	Multiplicity
miRCa-09	mirC-8853075	Eucgr.L03623	Eucgr.L03623.1	Osmotin 34	2.5	Cleavage	1
(Continuation)		Eucgr.D01898	Eucgr.D01898.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.D01904	Eucgr.D01904.1	Osmotin 34	2.5	Cleavage	1
		Eucgr.H00162	Eucgr.H00162.1	Phosphate transporter 1;7	3.0	Cleavage	1
		Eucgr.E03957	Eucgr.E03957.2	Indole-3-butyric acid response 1	3.0	Cleavage	1
		Eucgr.C02716	Eucgr.C02716.1	PPPDE putative thiol peptidase family protein	3.0	Cleavage	1
		Eucgr.J02626	Eucgr.J02626.1	PPPDE putative thiol peptidase family protein	3.0	Cleavage	1
		Eucgr.K02431	Eucgr.K02431.1	ARM repeat superfamily protein	3.0	Translation	1
		Eucgr.L02568	Eucgr.L02568.1	Osmotin 34	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.1	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.2	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.3	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.4	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.5	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
		Eucgr.J00040	Eucgr.J00040.6	UDP-glucuronic acid decarboxylase 3	3.0	Cleavage	1
miR171	ptc-miR171g|c|d	Eucgr.E01509	Eucgr.E01509.1	GRAS family transcription factor	0.5	Cleavage	1
		Eucgr.H01257	Eucgr.H01257	GRAS family transcription factor	0.5	Cleavage	1
	vvi-miR171a|i|f	Eucgr.E01509	Eucgr.E01509.1	GRAS family transcription factor	0.5	Cleavage	1
		Eucgr.H01257	Eucgr.H01257	GRAS family transcription factor	0.5	Cleavage	1

Legend: ptc – *Populus trichocarpa*; vvi – *Vitis vinifera*.

Supplementary Table S2. List of gene-specific primers (GSP) obtained using *Primer3 plus v2.3.3* software (Untergasser *et al.*, 2012) for RLM 5' -RACE validation of miRNA-directed cleavage target genes.

MiRNAs Candidate	Target Gene	Gene-specific primers Sequence (5'→ 3')	T_m (ºC)	GC (%)	Expected size (bp)
miR171	Eucgr.E01509	**GSP**: CCGAACTGGGGTCCGCTGCACAAGGC	73.4	69.2	-
		Nested GSP: GCCCATTGCCCACCATACCCAACATCA	69.2	55.6	324
	Eucgr.H01257	**GSP**: TGTCGGGCCAAGAGGAGAGCCGAGCC	73.4	69.2	-
		Nested GSP: GCCAAGAGGAGAGCCGAGCCGGA	70.3	69.6	176
mirCa-02	Eucgr.E01875	**GSP**: GCGACATGATTGCATCGCGAGACCTCGG	71.2	60.7	-
		Nested GSP: AGGATTCGTGCAACGTCGCTGGATCGA	70.1	55.6	269
	Eucgr.K00370	**GSP**: CATCGATGCACCCGATACATGAACGA	65.2	50.0	-
		Nested GSP: GCATCTACATCTCTTTCTCGACCACCCT	65.8	50.0	101
	Eucgr.B01226	**GSP**: TCTGGTTCCCAAGTTGCCTCAGGAAGT	67.6	51.9	-
		Nested GSP: CCAAGTTGCCTCAGGAAGTTTACATGGC	66.3	50.0	93
mirCa-04	Eucgr.B02279	**GSP**: GCCCAGACCTCCGGCTGTTGTTCGTCG	72.6	66.7	-
		Nested GSP: TCGCTGCATCTGGTTGGAAGCCCGA	70.4	60.0	218
	Eucgr.C01382	**GSP**: CCAAGAACAAGAGAAAAACCGAGGACGA	65.3	46.4	-
		Nested GSP: GAACAAGAGAAAAACCGAGGACGAA	61.7	44.4	329
	Eucgr.D01126	**GSP**: GCTCCAGCTTGGTTCCCACAGCCAGCA	72.6	63.0	-
		Nested GSP: CAGCTTGGTTCCCACAGCCAGCAGAAGA	70.4	57.1	171
mirCa-08	Eucgr.J02113	**GSP**: TGGGGGAGGAATGACAGGAGAGCGATGA	70.1	57.1	-
		Nested GSP: GGAGGAATGACAGGAGAGCGATGAG	64.5	56.0	85
	Eucgr.A01182	**GSP**: CGGCACTGGACTGCTGGTGGTAGTGC	71.1	65.4	-
		Nested GSP: AGGAAAGACCACTTCGGAGTCGTGCCT	69.2	55.6	179
	Eucgr. G02793	**GSP**: AGCTGTGGAGGAGGGTGAGGAGTGGCT	72.0	63.0	-
		Nested GSP: GTGGAGGAGGGTGAGGAGTGGCTGTGGA	72.2	64.3	68
mirCa-09	Eucgr.H03865	**GSP**: ACCAGGGCAGGTGAAAGTGCTGGTCTG	70.4	59.3	-
		Nested GSP: GGGCAGGTGAAAGTGCTGGTCTGATCA	68.6	55.6	101
	Eucgr.C02716	**GSP**: CCCTTCTGCGGTGATGATTCTAAGGCCC	68.7	57.1	-
		Nested GSP: TCTGCGGTGATGATTCTAAGGCCCT	65.7	52.0	133
	Eucgr.J00040	**GSP**: TGTTCCTGGAGCTTGCACAGTCAGCGGC	66.1	60.9	-
		Nested GSP: TTCCGCTACTCGCTTCCCCTCGTCA	69.2	60.0	501

Supplementary Table S3. List of the reference genes used for gene quantification by Quantitative Real-Time PCR assay (RT-qPCR; Cassan-Wang *et al.*, 2012).

Gene Code	Gene description	Gene model	Side and Primer Sequence
Helicase	RNA helicase family protein	Eucgr.G00345	**F:** AGGAGCAAACGCCTGTCAACTG
			R: ACATGCCCAATTGTGAACAGGAG
PP2A1	Protein phosphatase 2A-1	Eucgr.B03386	**F:** TCGAGCTTTGGACCGCATACAAG
			R: ACCACAAGAGGTCACACATTGGC
PP2A3	Protein phosphatase 2A subunit A3	Eucgr.B03031	**F:** CAGCGGCAAACAACTTGAAGCG
			R: ATTATGTGCTGCATTGCCCAGTC
PTB	Polypyrimidine tract-binding protein 1	Eucgr.J01358	**F:** TGTGGGAAGAGACGTTTGGAAGG
			R: AGACAAACCTGAGCCACTGAAGC
SAND-family	SAND family, trafficking protein Mon1	Eucgr.B02502	**F:** TTGATCCACTTGCGGACAAGGC
			R: TCACCCATTGACATACACGATTGC

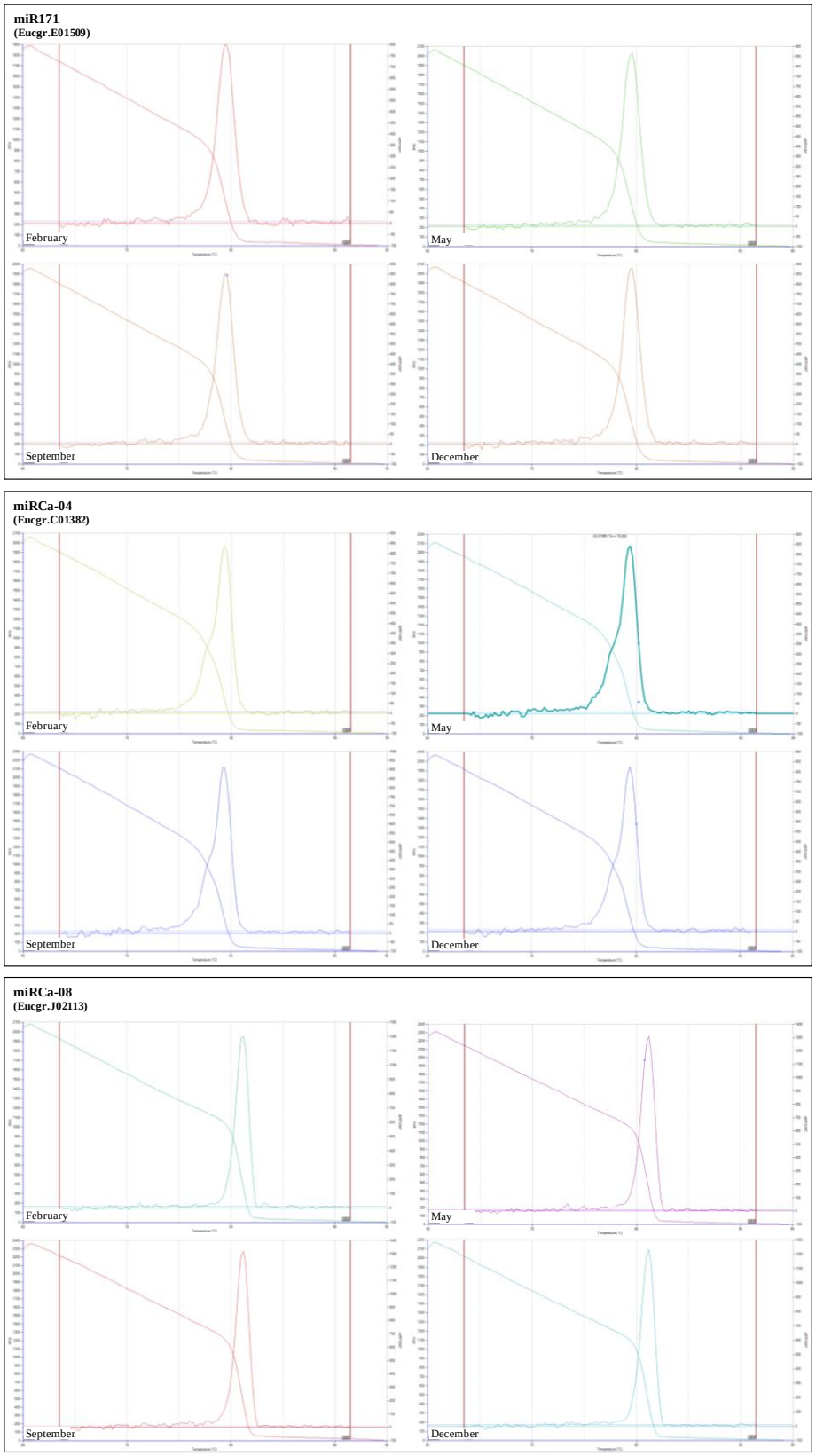
Supplementary Table S4. List of primers used for RT-qPCR assay.

MiRNAs Candidate	Transcript identifier	Target gene	Primers sequences (5'→ 3')	Length (bp)	T_m (ºC)	GC (%)
miR171	Egrandis_v1_0.004081m	Eucgr.E01509	**F:** TGGTGAGGCTATCGCTGTGAAC	22	62.9	54.5
			R: AGGAAGAATCGACGGGTGAGTTG	23	62.8	52.2
mirCa-04	Egrandis_v1_0.033753m	Eucgr.C01382	**F:** ACACGTACACGCACTCACACTC	22	63.3	54.5
			R: TTCCCACAGCTAGCTTGAACCG	22	63.2	54.5
mirCa-08	Egrandis_v1_0.010819m	Eucgr.J02113	**F:** AGGTCAAGGAGCTCGCAGTATAG	23	61.9	52.2
			R: GCAGTGTCGAATCCACCCAAATAC	24	62.6	50.0

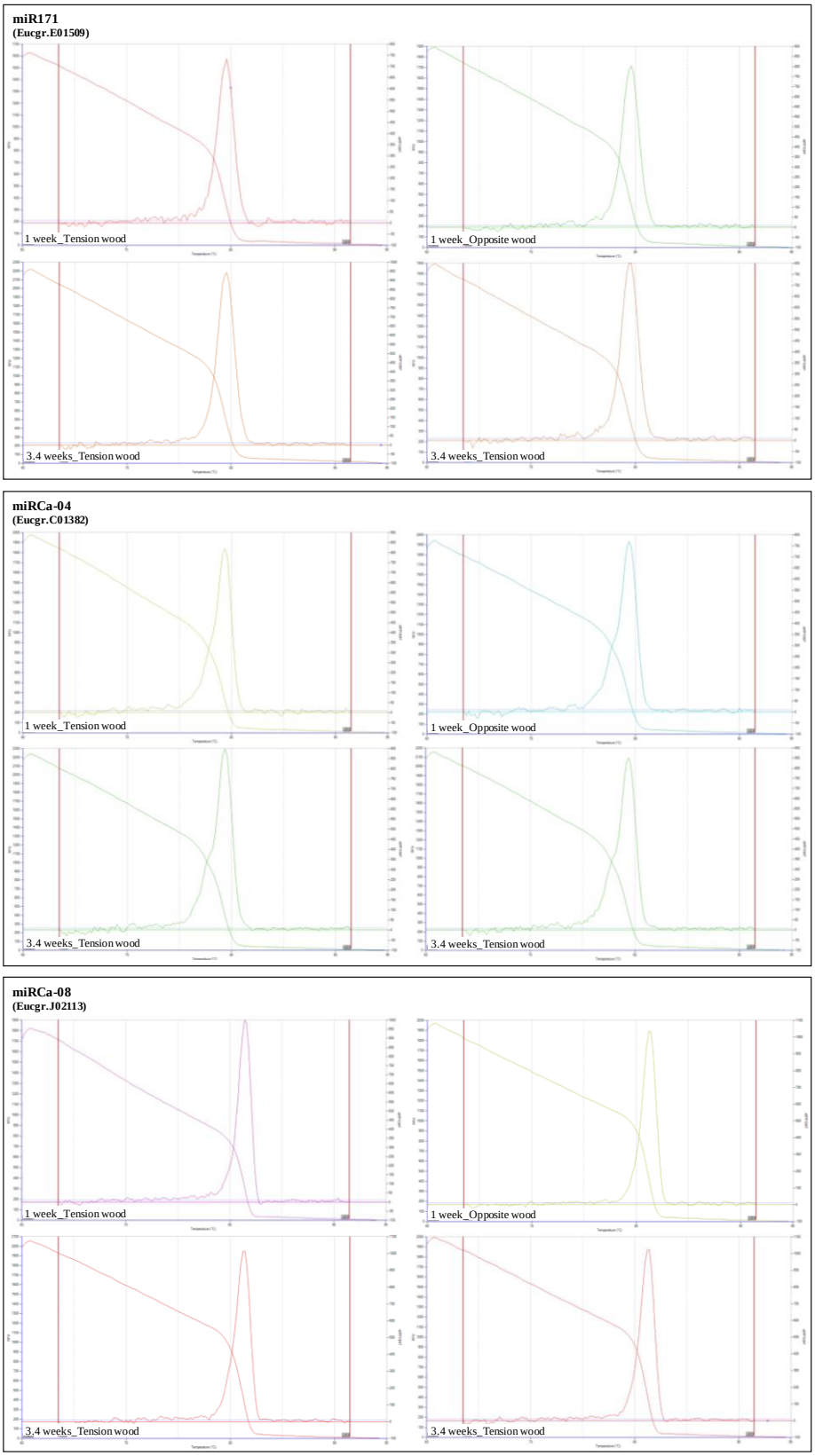
Supplementary Table S5. Statistical analysis (student's t-test) with 95% of confidence ($P<0.05$) for gene expression when compared the SW and RW_{2011} pools.

SW pools	**Eucgr.E01509**	**Eucgr.C01382**	**Eucgr.J02113**
Feb - May	0.3117	0.0486	0.9374
May - Sep	0.0017	0.0024	0.0014
Sep - Dec	0.0030	0.0002	0.0069
RW_{2010} pools			
1week_TW – 1week_OW	0.1961	0.0385	0.8116
1week_OW – 3.4weeks_TW	0.0317	0.9008	0.2967
3.4weeks_TW – 3.4weeks_OW	0.1612	0.2513	0.8427

Appendix II: Supplementary Figures



Supplementary Figure G1. Melting curves for seasonal wood (SW) pools using three target genes (Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113) by Quantitative Real-Time PCR (RT-qPCR) assay.



Supplementary Figure G2. Melting curves for reaction wood (RW_{2010}) pools using three target genes (Eucgr.E01509, Eucgr.C01382 and Eucgr.J02113) by Quantitative Real-Time PCR (RT-qPCR) assay.